现代数学基础丛书 115

矩阵理论与应用

（第二版）

陈公宁 编著



科学出版社
www.sciencep.com

(O-2843.0101)



销售分类建议：高等数学

定 价：56.00 元

内 容 简 介

本书系统介绍现代矩阵理论与应用的基本内容与预备知识。全书共分8章。主要内容包括:矩阵理论的基本知识,向量与矩阵的范数,矩阵函数,线性矩阵方程,矩阵与多项式的稳定性与惯性理论,矩阵的广义逆,矩阵特征值的定位与扰动,非负矩阵的 Perron-Frobenius 理论及其推广,以及 *M*-矩阵理论及其在数理经济学的投入-产出模型分析中的应用等。内容丰富、翔实,并配备有大量的练习题。

本书可作为高等院校数学系高年级本科生、研究生,特别是计算数学与应用数学专业的研究生教材,也可供相关工程技术专业的教师、科研人员阅读参考.

图书在版编目(CIP)数据

矩阵理论与应用/陈公宁编著. —2 版. —北京:科学出版社, 2007
(现代数学基础丛书;115)

ISBN 978-7-03-019531-9

Ⅰ. 矩… Ⅱ. 陈… Ⅲ. 矩阵-理论 Ⅳ. O151. 21

中国版本图书馆 CIP 数据核字(2007)第 118542 号

责任编辑:张 扬 贾瑞娜/责任校对:邹慧卿

责任印制:赵德静/封面设计:陈 敬

科 学 出 版 社 出版
北京东黄城根北街 16 号
邮政编码:100717
http://www.sciencep.com
新 蕾 印 刷 厂 印刷
科学出版社发行 各地新华书店经销

*

1990 年 9 月第 一 版 开本:B5(720×1000)
2007 年 8 月第 二 版 印张:23 1/4
2007 年 8 月第一次印刷 字数:438 000
印数:1—3 000

定价:56.00 元

《现代数学基础丛书》序

对于数学研究与培养青年数学人才而言，书籍与期刊起着特殊重要的作用. 许多成就卓越的数学家在青年时代都曾钻研或参考过一些优秀书籍，从中汲取营养，获得教益.

20 世纪 70 年代后期，我国的数学研究与数学书刊的出版由于文化大革命的浩劫已经被破坏与中断了十余年，而在这期间国际上数学研究却在迅猛地发展着. 1978 年以后，我国青年学子重新获得了学习、钻研与深造的机会. 当时他们的参考书籍大多还是 50 年代甚至更早期的著述. 据此，科学出版社陆续推出了多套数学丛书，其中《纯粹数学与应用数学专著》丛书与《现代数学基础丛书》更为突出，前者出版约 40 卷，后者则逾 80 卷. 它们质量甚高，影响颇大，对我国数学研究、交流与人才培养发挥了显著效用.

《现代数学基础丛书》的宗旨是面向大学数学专业的高年级学生、研究生以及青年学者，针对一些重要的数学领域与研究方向，作较系统的介绍. 既注意该领域的基础知识，又反映其新发展，力求深入浅出，简明扼要，注重创新.

近年来，数学在各门科学、高新技术、经济、管理等方面取得了更加广泛与深入的应用，还形成了一些交叉学科. 我们希望这套丛书的内容由基础数学拓展到应用数学、计算数学以及数学交叉学科的各个领域.

这套丛书得到了许多数学家长期的大力支持，编辑人员也为其付出了艰辛的劳动. 它获得了广大读者的喜爱. 我们诚挚地希望大家更加关心与支持它的发展，使它越办越好，为我国数学研究与教育水平的进一步提高作出贡献.

杨 乐

2003 年 8 月

再版序言

这一版有较大的变动。广大读者对本书初版的意见与建议，我都尽量地予以考虑与接纳。最重要的变动是调整并压缩了全书的内容。具体如下：①删去1.3.2小节，而把其中1.3.1小节改成4.1.1小节；②去掉2.4节，它的某些结果作为练习题见于其他章节的习题中；③删去6.4节矩阵的解析扰动的内容；④删去9.1节，而把9.2节改成的最后一节；⑤在余下的各章节中还有一些改变，主要是简化了许多基本定理、推论的证明，并将一些非基本但又带有繁重证明的结果改做练习题。同时也增添了一些基本性的练习题。此外，对初版中的明显错误与疏漏，进行校正与补遗。

这一版仍保留初版中各章的相对独立性。读者不必按章节的次序，可以比较随意地（或只需做少量补充便可）阅读其中的某一部分。根据我们的教学经验，书中的第二章（可删去2.3节）到第五章与第六章到第八章可分别作为大学数学系或有关专业高年级学生或研究生的一学期选修教材。如有可能，在讲授时可以指导学生讨论书中某一专题的结果，并完成相关的练习题中较难的部分。

多年来许多读者，尤其是我的学生们与同仁对本书的不足提出了许多有益的意见与建议，从而极大地帮助了本书的修订。北京师范大学数学科学学院的领导与许多同志在促使本书早日再版过程中做了许多努力。科学出版社的张扬编辑在本书再版编辑工作中给予了很大的帮助与支持。在此恳切地向他们表示感谢。同时也感谢我的夫人郑欣女士，我的女儿一家与其他亲属在本书修订过程中的帮助与关怀。

恳请读者对再版的缺点仍不吝指教，并对其内容与材料安排等方面提出意见与建议，如果读者肯惠予书面回馈，那就更为感激了。

陈公宁
2007年1月于北京师大乐育楼区

初版序言

全国高等学校计算数学与逻辑学教材编审组的近几次年会，特别是1987年广州年会商定为综合性大学计算数学专业研究生编写“矩阵理论与应用”教材，并撰写了编写大纲。

众所周知，矩阵论在历史上至少可追溯到Sylvester与Cayley，特别是Cayley 1858年的工作。近代数学的一些学科，如代数结构理论与泛函分析可以在矩阵论中寻到它们的根源。另一方面，作为一种基本工具，矩阵论在应用数学与工程技术学科，如微分方程、概率统计、最优化、运筹学、计算数学、控制论与系统理论等有着广泛的应用。这些学科的发展反过来又极大地促进了矩阵论的发展。根据这些特点，并参考近年来出版的多种矩阵论著作以及根据我们多年来矩阵论教学的实践经验，在本书选材中，力求选取矩阵论里既具有基本理论意义又具有重要应用价值的一部分古典与现代内容以满足计算数学专业与相近专业，如数学、应用数学、工程控制、系统科学等不同类别与不同层次读者的需要。我们假定读者具有一般微积分与线性代数的知识背景(少数几处还要求一点复分析的基础知识). 本书可作为数学、计算数学、应用数学与某些工程技术学科本科高年级学生或研究生的教材或参考书。

按照大纲的设想，讲授本书基本内容约需80学时，余下内容为选学材料，对它们不提出基本要求，教师可酌情取舍。对于熟悉线性代数的读者，建议可在简单复习第一章基本知识(特别是§2和§3)后直接进入后面的章节。第二章介绍向量与矩阵范数，其中§1.1～§3.2是基本的，§3.3～§4可作为选学内容。第三章研究矩阵函数，为本书重点内容之一，除了§4以外应力求讲授余下的大部分内容。第四章讨论线性矩阵方程与惯性理论，以§1.1～§2.2为基本内容，§3可以略去(但对控制论专业的读者来说，第四章全面内容是必要的)。第五章讲授矩阵的广义逆知识，其中§2.3可以略去。第六章介绍特征值的定位与扰动，其中§4可以略去(但对§3内容稍有点影响)。最后三章研究非负矩阵的Perron-Frobenius理论与M-矩阵理论及其在求解线性方程组迭代方法与数理经济学Leontief模型分析中的应用。其中，第七章§2.2，第八章§2.2～2.3可以删去或略讲，第九章两节可取其一(其中§1.3可以不讲或略讲)。这三章基本内容构成本书另一个重点。

为便于引证，本书将所有出现在每一小节中的定义、引理、定理、命题与例子按前后顺序统一排列。每一小节后配有习题，全书备有484道习题，其中不少是正文

的延伸与深化，读者应尽量完成其中较为基本的部分，并能阅读余下的部分。每章后面罗列参考文献。如有可能应进一步收集与钻研有关文献，特别是 AMS 的 Mathematical Rveiews 2D-15 部分的那些资料。

中国科学院计算中心孙继广研究员审阅了本书，提出了不少宝贵的修改意见与建议，谨致以衷心的感谢。我也感谢牛少彰、许松、梁景伟与杨正宏等同志多年来的合作与帮助。

由于水平所限，错漏之处在所难免，欢迎读者批评指正。

陈公宁

1989 年 6 月于北京师范大学

目　　录

第一章　矩阵理论的基本知识

本章温习与介绍矩阵论的有关概念与基本结论，作为后面各章的预备知识. 其中一些结果可以在普通的线性代数教程(如文献[7],[8],[10]～[12])中找到，对此我们仅简单回顾一下有关结果，指明思路或辅以实例说明. 特别地，假定读者熟悉域上线性空间、子空间、线性相关与线性无关、基、空间维数等基本概念与结果. 本章也包含一些普通读者可能较生疏的内容，例如，Hermite 正定与正半定矩阵的性质、某些矩阵的谱定理、Schur 余量及其应用等，对这一部分内容，进行比较详细的讨论. 本章内容主要取自文献[1]～[6],[9].

1.1　矩阵与线性变换

设 $\boldsymbol{F}$ 表示某个域. $\boldsymbol{F}$ 内的 mn 个元素 $a_{11},a_{12},a_{13},\cdots,a_{21},\cdots,a_{m1},\cdots,a_{mn}$ 的如下矩形阵列：

$$A=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

称为一个有 m 行与 n 列的 $m\times n$ **矩阵**，通常简记为

$$A=(a_{ij})_{i,j=1}^{m,n}\quad\text{或}\quad A=(a_{ij})\in \boldsymbol{M}_{m,n}(\boldsymbol{F}),$$

这里，$\boldsymbol{M}_{m,n}(\boldsymbol{F})$ 表示这样 $m\times n$ 矩阵的集合. $m\times n$ 叫作矩阵 A 的**大小或维数**；a_{ij} 叫作矩阵 A 在位置 (i,j) 处(即第 i 行与第 j 列交叉处)的**元素**. $m\times n$ 矩阵 A 的第 i 行元素 $a_{i1},a_{i2},\cdots,a_{in}$ 的全体记为

$$A_{(i)}=[a_{i1}\quad a_{i2}\quad \cdots\quad a_{in}],\quad i=1,2,\cdots,m.$$

每个 $A_{(i)}$ 本身可看成一个 $1\times n$ 矩阵. 同样地，$m\times n$ 矩阵 A 的第 j 列元素 a_{1j}，$a_{2j},\cdots,a_{mj}$ 的全体记为

$$A^{(j)}=\begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix},\quad j=1,2,\cdots,n.$$

每个 $A^{(j)}$ 本身为一个 $m\times 1$ 矩阵.

当域 $\boldsymbol{F}$ 取实数域$\mathbb{R}$(或复数域$\mathbb{C}$)时，$\boldsymbol{M}_{m,n}(\mathbb{R})$(或 $\boldsymbol{M}_{m,n}(\mathbb{C})$)便是 $m\times n$ 实矩阵

(或复矩阵)的集合. 更一般地,矩阵元素可以取为关于λ的多项式或函数等. 但是,若无特别声明,本书内$\boldsymbol{F}$总是表示$\mathbb{R}$或$\mathbb{C}$. 若$m=n$,则称A为**n阶矩阵**或**n阶方阵**,记为$A=(a_{ij})_{i,j=1}^{n}$. 通常用记号$\boldsymbol{M}_n(\boldsymbol{F})$简记$\boldsymbol{M}_{n,n}(\boldsymbol{F})$. 若$m=1$或$n=1$,则$\boldsymbol{M}_{1,n}(\boldsymbol{F})$便为$n$维行向量的集合,$\boldsymbol{F}^m\equiv\boldsymbol{M}_{m,1}(\boldsymbol{F})$便为$m$维列向量的集合(在本书中用符号"$\equiv$"表示某关系式可作为定义).

两个矩阵$A=(a_{ij}),B=(b_{ij})\in\boldsymbol{M}_{m,n}(\boldsymbol{F})$叫作**相等的**,并记为$A=B$,假如它们对应的元素均相等,即有$a_{ij}=b_{ij},i=1,\cdots,m;j=1,\cdots,n$.

今后,一般用大写英文字母表示矩阵,小写英文字母(带双下标)或数表示矩阵元素;用小写黑体英文字母表示$\boldsymbol{F}^m$中向量,用小写英文字母(带单下标)或数表示该向量的元素,例如,$\boldsymbol{x}=(x_1,x_2,\cdots,x_m)^{\mathrm{T}},\boldsymbol{y}=(y_1,y_2,\cdots,y_m)^{\mathrm{T}},\boldsymbol{e}=(1,1,\cdots,1)^{\mathrm{T}}$,等等,其中上标T表示转置. 特殊地,$O$表示元素皆为零的零矩阵,若需要指明大小,可记为$O_{m,n},O_n$等. $\boldsymbol{F}^m$中零向量记为$\boldsymbol{x}=\boldsymbol{0}$,从上下文,一般地容易判别零向量的维数.

1.1.1 矩阵与行列式,特征值与特征向量

为了能够处理矩阵,需要一些基本运算.

设$A=(a_{ij}),B=(b_{ij})\in\boldsymbol{M}_{m,n}(\boldsymbol{F})$. 定义$C=(c_{ij})\in\boldsymbol{M}_{m,n}(\boldsymbol{F})$为$A$与$B$**之和**,并且记为$C=A+B$,假如

$$c_{ij}=a_{ij}+b_{ij},\qquad i=1,\cdots,m;j=1,\cdots,n. \tag{1.1}$$

对$\alpha\in\boldsymbol{F}$,定义$\alpha A=(\alpha a_{ij})\in\boldsymbol{M}_{m,n}(\boldsymbol{F})$为**数$\boldsymbol{\alpha}$与矩阵$\boldsymbol{A}$之乘积**. 在上述加法与数乘运算之下,$\boldsymbol{M}_{m,n}(\boldsymbol{F})$组成域$\boldsymbol{F}$上的线性向量空间. 这是因为若$A,B,C\in\boldsymbol{M}_{m,n}(\boldsymbol{F})$. $\alpha,\beta\in\boldsymbol{F}$,则有

$$A+B=B+A,$$
$$(A+B)+C=A+(B+C),$$

存在$m\times n$零矩阵O,使得$A+O=A$. 对每一个$A\in\boldsymbol{M}_{m,n}(\boldsymbol{F})$,存在$-A\equiv(-1)A$使得$A+(-A)=O$,

$$\alpha(\beta A)=(\alpha\beta)A,$$
$$\alpha(A+B)=\alpha A+\alpha B,$$
$$(\alpha+\beta)A=\alpha A+\beta A,$$
$$1A=A.$$

矩阵乘法可以定义为自$\boldsymbol{M}_{m,n}(\boldsymbol{F})\times\boldsymbol{M}_{n,q}(\boldsymbol{F})$到$\boldsymbol{M}_{m,q}(\boldsymbol{F})$内的如下二元运算:若$A=(a_{ij})\in\boldsymbol{M}_{m,n}(\boldsymbol{F}),B=(b_{ij})\in\boldsymbol{M}_{n,q}(\boldsymbol{F})$,则

$$AB=C=(c_{ij})\in\boldsymbol{M}_{m,q}(\boldsymbol{F}),$$

其中,

$$c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}, \quad i = 1,\cdots,m; j = 1,\cdots,q. \tag{2}$$

矩阵 $C=AB$ 称为 **A 与 B 的乘积**. 注意,矩阵乘积 AB 的定义要求 A 的列数等于 B 的行数. 容易验证:假如下列各式有定义,则有

$$\begin{aligned} (AB)D &= A(BD), \\ A(B+D) &= AB+AD, \\ (B+D)A &= BA+DA, \\ (\alpha A)B &= A(\alpha B) = \alpha(AB), \alpha \in \boldsymbol{F}. \end{aligned} \tag{3}$$

注意,矩阵乘法除了一阶情形外不满足交换律,即 $AB=BA$ 一般不成立.

当 $m=n$ 时由(3)式看出,按前述三种运算,$\boldsymbol{M}_n(\boldsymbol{F})$不仅组成 $\boldsymbol{F}$ 上的线性空间,且为 $\boldsymbol{F}$ 上的**线性结合代数**. 这里 n 阶单位矩阵 $I_n=(\delta_{ij})$(δ_{ij} 表示 Kronecker 符号)为此代数中的乘法单位元,即有 $AI_n=I_nA=A$, $\forall A\in\boldsymbol{M}_n(\boldsymbol{F})$.

若 $A\in\boldsymbol{M}_n(\boldsymbol{F})$,则可考虑 A 的非负整数幂:

$$A^p \equiv \underbrace{AA\cdots A}_{p\text{次}}(p \geqslant 1), \qquad A^0 \equiv I_n.$$

除了零矩阵与单位矩阵以外,我们今后会经常遇到如下几种特殊结构的 n 阶矩阵 $A=(a_{ij})$:

对角矩阵,记为 $\mathrm{diag}[a_{11},a_{22},\cdots,a_{nn}]$: $a_{ij}=0, i,j=1,\cdots,n; i\neq j$. 单位矩阵 I_n 显然为特殊的对角矩阵. 若所有 $a_{ii}\geqslant 0$ ($a_{ii}>0$),则称 $\mathrm{diag}[a_{11},a_{22},\cdots,a_{nn}]$为 n 阶**非负(正)对角矩阵**.

上三角形矩阵: $a_{ij}=0$, $\forall i>j$.

下三角形矩阵: $a_{ij}=0$, $\forall i<j$.

幂等矩阵: $A^2=A$.

幂零矩阵: $A^r=O$,其中 r 为某正整数.

对称矩阵: $a_{ij}=a_{ji}, i,j=1,\cdots,n$.

Hermite(或自共轭)矩阵: $a_{ij}=\bar{a}_{ji}, i,j=1,\cdots,n$,其中,$\bar{a}_{ji}$表示 a_{ji} 的共轭复数.

若将元素 $a_{11},a_{22},\cdots,a_{nn}$定义为$A=(a_{ij})\in\boldsymbol{M}_n(\boldsymbol{F})$的**主对角线**,则 n 阶对角矩阵便为非主对角线元素皆为零的 n 阶矩阵,上(下)三角形矩阵便为主对角线以下(以上)元素皆为零的 n 阶矩阵. 若用右上标"T"表示矩阵的**转置**,即对 $A=(a_{ij})_{i,j=1}^{m,n}\in\boldsymbol{M}_{m,n}(\boldsymbol{F})$有 $A^{\mathrm{T}}\in\boldsymbol{M}_{n,m}(\boldsymbol{F})$,它的$(i,j)$元素等于 a_{ji},用右上标"$*$"表示矩阵的**共轭转置**,即对 $A=(a_{ij})_{i,j=1}^{m,n}\in\boldsymbol{M}_{m,n}(\boldsymbol{F})$有 $A^*\in\boldsymbol{M}_{n,m}(\boldsymbol{F})$,它的$(i,j)$元素等于 $\bar{a}_{ji}$,那么,当 $A=(a_{ij})\in\boldsymbol{M}_n(\boldsymbol{F})$时,对称矩阵可由 $A=A^{\mathrm{T}}$ 来表征,Hermite 矩阵可由$A=A^*$ 来表征. 显然,若 AB 有定义,则有

$$(AB)^{\mathrm{T}} = B^{\mathrm{T}}A^{\mathrm{T}}, \quad (AB)^* = B^*A^*, \quad (A^{\mathrm{T}})^{\mathrm{T}} = A, \quad (A^*)^* = A.$$

设 $1\leqslant k\leqslant n$,我们用 $Q_{k,n}$表示集合 $\mathcal{N}=\{1,2,\cdots,n\}$中所有 k 个整数组成的严格

上升序列的全体，即

$$Q_{k,n} = \{\alpha = (i_1, i_2, \cdots, i_k): 1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant n\}. \tag{4}$$

显然，$Q_{k,n}$中有$\binom{n}{k}=\frac{n!}{k!\ (n-k)!}$个序列. 例如，$Q_{2,3}=\{(1,2),(1,3),(2,3)\}$. 应用这个记号，我们可以比较方便地表示矩阵$A=(a_{ij})\in \boldsymbol{M}_{m,n}(\boldsymbol{F})$的**子矩阵**. 假定$k$与$r$为满足$1\leqslant k\leqslant m$与$1\leqslant r\leqslant n$的正整数，且$\alpha=(i_1,i_2,\cdots,i_k)\in Q_{k,m}$，$\beta=(j_1,j_2,\cdots,j_r)\in Q_{r,n}$，则$(s,t)$元素为$a_{i_s j_t}(s=1,\cdots,k;t=1,\cdots,r)$的矩阵$B\in \boldsymbol{M}_{k,r}(\boldsymbol{F})$叫作(行序在$\alpha$，列序在$\beta$)**$A$的子(矩)阵**，记为

$$B = A[\alpha \mid \beta]. \tag{5}$$

若$A=(a_{ij})\in \boldsymbol{M}_n(\boldsymbol{F})$，则

$$A[\alpha] \equiv A[\alpha \mid \alpha], \alpha \in Q_{k,n} \tag{6}$$

叫作A的(行列序在α)k阶**主子阵**. 为了今后方便，我们也用记号$A[\alpha|\beta)$，$A(\alpha|\beta]$与$A(\alpha|\beta)$分别表示A的行序在α而列序不在β的子阵，A的行序不在α而列序在β的子阵，A的行序不在α且列序不在β的子阵. 例如，$A=(a_{ij})\in \boldsymbol{M}_{5,6}(\boldsymbol{F})$，$k=2$，$r=3$，$\alpha=(1,3)$，$\beta=(2,3,4)$，这时有

$$A[\alpha \mid \beta] = \begin{bmatrix} a_{12} & a_{13} & a_{14} \\ a_{32} & a_{33} & a_{34} \end{bmatrix} \in \boldsymbol{M}_{2,3}(\boldsymbol{F}),$$

$$A[\alpha \mid \beta) = \begin{bmatrix} a_{11} & a_{15} & a_{16} \\ a_{31} & a_{35} & a_{36} \end{bmatrix} \in \boldsymbol{M}_{2,3}(\boldsymbol{F}),$$

$$A(\alpha \mid \beta] = \begin{bmatrix} a_{22} & a_{23} & a_{24} \\ a_{42} & a_{43} & a_{44} \\ a_{52} & a_{53} & a_{54} \end{bmatrix} \in \boldsymbol{M}_{3,3}(\boldsymbol{F}),$$

$$A(\alpha \mid \beta) = \begin{bmatrix} a_{21} & a_{25} & a_{26} \\ a_{41} & a_{45} & a_{46} \\ a_{51} & a_{55} & a_{56} \end{bmatrix} \in \boldsymbol{M}_{3,3}(\boldsymbol{F}),$$

$$A[\alpha \mid \alpha] = \begin{bmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{bmatrix} \in \boldsymbol{M}_2(\boldsymbol{F}).$$

存在一种特殊的方式将一个矩阵分为若干个子阵，即用贯穿整个矩阵的一组水平线与垂直线将矩阵划分为若干块，每块皆为原先矩阵的子阵. 例如，矩阵$A=(a_{ij})\in \boldsymbol{M}_{3,4}(\boldsymbol{F})$可以划分为

$$A = \left[\begin{array}{c:cc:c} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ \hdashline a_{31} & a_{32} & a_{33} & a_{34} \end{array}\right] \quad \text{或} \quad A = \left[\begin{array}{cc:cc} a_{11} & a_{12} & a_{13} & a_{14} \\ \hdashline a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{array}\right],$$

等等. 按前一种划分，A可写成如下块矩阵形式

$$A=\begin{bmatrix}A_{11} & A_{12} & A_{13}\\ A_{21} & A_{22} & A_{23}\end{bmatrix},\tag{7}$$

其中，

$$A_{11}=\begin{bmatrix}a_{11}\\ a_{21}\end{bmatrix},\quad A_{12}=\begin{bmatrix}a_{12} & a_{13}\\ a_{22} & a_{23}\end{bmatrix},\quad A_{13}=\begin{bmatrix}a_{14}\\ a_{24}\end{bmatrix},$$

$$A_{21}=[a_{31}],\quad A_{22}=[a_{32}\quad a_{33}],\quad A_{23}=[a_{34}].$$

这里某些块，如 A_{21} 与 A_{23} 可以为一阶矩阵，即 $\boldsymbol{F}$ 中的数量. 按后一种划分，A 可写成

$$A=\begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix},$$

其中，

$$A_{11}=[a_{11}\quad a_{12}],\quad A_{12}=[a_{13}\quad a_{14}],$$

$$A_{21}=\begin{bmatrix}a_{21} & a_{22}\\ a_{31} & a_{32}\end{bmatrix},\quad A_{22}=\begin{bmatrix}a_{23} & a_{24}\\ a_{33} & a_{34}\end{bmatrix}.$$

这样划分成块的矩阵 A 叫作**分块矩阵**，它的块可视为分块矩阵的元素，记为

$$A=[A_{ij}]_{i,j=1}^{k,s},\tag{8}$$

它表示 A 为 k 块行与 s 块列的分块矩阵. 类似于含有数量元素的矩阵，我们可以定义分块方阵，假若(8)式中 $k=s$；分块对角矩阵，假若 $A_{ij}=O,\ \forall i\neq j, i,j=1,\cdots,k=s$，分块上(下)三角形矩阵，假若 $A_{ij}=O,\ \forall i>j, i,j=1,\cdots,k=s$ ($A_{ij}=O,\ \forall i<j, i,j=1,\cdots,k=s$). 分块对角矩阵可记为

$$\operatorname{diag}[A_{11},A_{22},\cdots,A_{kk}].\tag{9}$$

对于方阵 $A=(a_{ij})$，今后我们常遇到的是对角线上各块都是 A 的主子阵的所谓**对称分块矩阵**. 例如，

$$A=\left[\begin{array}{c:cc}a_{11} & a_{12} & a_{13}\\ \hdashline a_{21} & a_{22} & a_{23}\\ a_{31} & a_{32} & a_{33}\end{array}\right].$$

对于分块矩阵我们也可以引入加法、数乘与乘法运算. 假如 A 与 B 为有相同大小且有一致的分块形式的分块：

$$A=[A_{ij}]_{i,j=1}^{k,s},\quad B=[B_{ij}]_{i,j=1}^{k,s},$$

则显然有

$$A+B=[A_{ij}+B_{ij}]_{i,j=1}^{k,s},\quad \alpha A=[\alpha A_{ij}]_{i,j=1}^{k,s},\tag{10}$$

这里，$\alpha\in\boldsymbol{F}$. 假如 $A\in\boldsymbol{M}_{m,l}(\boldsymbol{F})$ 与 $B\in\boldsymbol{M}_{l,n}(\boldsymbol{F})$ 分块为

$$A=[A_{ik}]_{i,k=1}^{r,s},\quad B=[B_{kj}]_{k,j=1}^{s,t},$$

其中，A_{ik} 大小为 $m_i\times l_k$，B_{kj} 大小为 $l_k\times n_j$，则容易验证

$$AB = \Big[\sum_{k=1}^{s} A_{ik}B_{kj}\Big]_{i,j=1}^{r,t}. \tag{11}$$

因此，当对分块矩阵进行加法、数乘与乘法运算时，其规则基本上与将块换成数后矩阵相应的运算是一致的，当然，这时要注意到块之间的"相容性".

例 1　(1) 设

$$A = \begin{bmatrix} M & \boldsymbol{u} \\ \boldsymbol{v}^{\mathrm{T}} & a \end{bmatrix},\quad B = \begin{bmatrix} Q & \boldsymbol{y} \\ \boldsymbol{x}^{\mathrm{T}} & b \end{bmatrix},$$

其中，M 与 Q 为 $n-1$ 阶矩阵；a 与 b 为数量；$\boldsymbol{x},\boldsymbol{y},\boldsymbol{u},\boldsymbol{v}\in \boldsymbol{F}^{n-1}$. 则

$$A+B = \begin{bmatrix} M+Q & \boldsymbol{u}+\boldsymbol{y} \\ \boldsymbol{v}^{\mathrm{T}}+\boldsymbol{x}^{\mathrm{T}} & a+b \end{bmatrix},\quad \alpha A = \begin{bmatrix} \alpha M & \alpha \boldsymbol{u} \\ \alpha \boldsymbol{v}^{\mathrm{T}} & \alpha a \end{bmatrix}, \alpha \in \boldsymbol{F},$$

$$AB = \begin{bmatrix} MQ+\boldsymbol{u}\boldsymbol{x}^{\mathrm{T}} & M\boldsymbol{y}+\boldsymbol{u}b \\ \boldsymbol{v}^{\mathrm{T}}Q+a\boldsymbol{x}^{\mathrm{T}} & \boldsymbol{v}^{\mathrm{T}}\boldsymbol{y}+ab \end{bmatrix}.$$

(2)

$$\left[\begin{array}{cc:cc} 1 & 0 & a_{11} & a_{12} \\ 0 & 1 & a_{21} & a_{22} \\ \hdashline 0 & 0 & b_{11} & b_{12} \\ 0 & 0 & b_{21} & b_{22} \end{array}\right]\left[\begin{array}{cc:cc} c_{11} & c_{12} & -a_{11} & -a_{12} \\ c_{21} & c_{22} & -a_{21} & -a_{22} \\ \hdashline 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right] = \begin{bmatrix} I_2 & A \\ O & B \end{bmatrix}\begin{bmatrix} C & -A \\ O & I_2 \end{bmatrix}$$

$$= \begin{bmatrix} I_2C+AO & -I_2A+AI_2 \\ OC+BO & -OA+BI_2 \end{bmatrix} = \begin{bmatrix} C & O \\ O & B \end{bmatrix}. \tag{12}$$

(3) 设 A 为如下对称分块矩阵：

$$A = \begin{bmatrix} A_{11} & A_{12} \\ O & I_n \end{bmatrix},\quad A_{11} \in M_m(\boldsymbol{F}).$$

则对任意正整数 p，

$$A^p = \begin{bmatrix} A_{11}^p & q(A_{11})A_{12} \\ O & I_n \end{bmatrix},$$

式中，$q(A_{11}) = A_{11}^{p-1} + A_{11}^{p-2} + \cdots + I_m$.

(4) 对于对称分块矩阵来说，两个可乘的分块上三角形矩阵之积仍为分块上三角形矩阵，两个可乘的分块下三角形矩阵之积仍为分块下三角形矩阵，两个可乘的分块对角矩阵之积仍为分块对角矩阵.　□

与方阵紧密相关的概念是**行列式与非奇异矩阵**. 对 $A=(a_{ij})\in \boldsymbol{M}_n(\boldsymbol{F})$，$A$ 的**行列式**（记为 $\det A$）定义如下：

$$\det A = \sum_{\sigma\in S_n} \mathrm{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}, \tag{13}$$

其中，S_n 为集合 $\mathscr{N}=\{1,2,\cdots,n\}$ 上所有置换 σ 组成的 n 阶对称群，且若 $\sigma\in S_n$ 为偶置换，$\mathrm{sgn}(\sigma)=1$，否则，$\mathrm{sgn}(\sigma)=-1$. 因此，$\det A$ 是 $n!$ 个诸如 $\pm a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}$

乘积之和.

按(13)式,我们有

$$\det I_n = 1, \quad \det A^{\mathrm{T}} = \det A, \quad \det A^* = \overline{\det A},$$
$$\det(\alpha A) = \alpha^n \det A, \quad \forall \alpha \in \boldsymbol{F}.$$

行列式还有许多重要性质,罗列一部分如下.

定理 2(Laplace)　设 $A=(a_{ij})\in \boldsymbol{M}_n(\boldsymbol{F})$,$1\leqslant r\leqslant n$,$\alpha\in Q_{r,n}$. 对任意 $\beta=(i_1,i_2,\cdots,i_r)\in Q_{r,n}$,令

$$s(\beta) = \sum_{t=1}^{r} i_t.$$

则有关于按 α 中行与列的展开式

$$\begin{aligned} \det A &= (-1)^{s(\alpha)} \sum_{\beta} (-1)^{s(\beta)} \det A[\alpha \mid \beta] \det A(\alpha \mid \beta), \\ \det A &= (-1)^{s(\alpha)} \sum_{\beta} (-1)^{s(\beta)} \det A[\beta \mid \alpha] \det A(\beta \mid \alpha), \end{aligned} \tag{14}$$

式中,β 取遍集合 $Q_{r,n}$.

特殊地,取 $r=1$,从(14)式推出

$$\det A = \sum_{j=1}^{n} a_{ij} A_{ij} = \sum_{i=1}^{n} a_{ij} A_{ij}, \tag{15}$$

这里 $A_{ij}=(-1)^{i+j}\det A(i|j)$即是 a_{ij} 的**代数余子式**. (15)式给出 A 的行列式 $\det A$ 分别关于第 i 行与第 j 列的展开. 因此,$\det A$ 可视为 A 的第 i 个行向量的线性函数,确切地说,若 $A_{(i)}=\alpha B_{(i)}+\beta C_{(i)}$,$\alpha,\beta\in F$,则

$$\det\begin{bmatrix} A_{(1)} \\ \vdots \\ \alpha B_{(i)} + \beta C_{(i)} \\ \vdots \\ A_{(n)} \end{bmatrix} = \alpha \det\begin{bmatrix} A_{(1)} \\ \vdots \\ B_{(i)} \\ \vdots \\ A_{(n)} \end{bmatrix} + \beta \det\begin{bmatrix} A_{(1)} \\ \vdots \\ C_{(i)} \\ \vdots \\ A_{(n)} \end{bmatrix}.$$

对于列也有类似的结果. 此外,按(15)式中关于行列展开结果,若 A 的某行(列)为零向量,则 $\det A=0$;按定理 2,具有相同两行(列)的方阵 A 有 $\det A=0$. 因此,我们有

$$\begin{aligned} &\sum_{k=1}^{n} a_{jk} A_{ik} = 0; \quad i,j \in \mathcal{N}; i \neq j, \\ &\sum_{i=1}^{n} a_{il} A_{ik} = 0, \quad k,l \in \mathcal{N}; k \neq l. \end{aligned} \tag{16}$$

应用 $\det A$ 为 A 的某个列向量的线性函数的事实与定理 2 关于列展开结果,我们可得:若 $A,B\in \boldsymbol{M}_n(\boldsymbol{F})$,则

$$\det(A+B) = \sum_{\alpha,\beta} \det A[\alpha \mid \beta] \operatorname{codet} B[\alpha \mid \beta], \tag{17}$$

其中，α 与 β 取遍 $\mathcal{N}$ 的不同子集，且 α 与 β 的元素个数相等，而 $\mathrm{codet}B[\alpha|\beta]$ 为 $\det B[\alpha|\beta]$ 的代数余子式，即

$$\mathrm{codet}B[\alpha \mid \beta] = (-1)^{s(\alpha)+s(\beta)}\det B[\alpha \mid \beta]. \tag{18}$$

关于两个矩阵乘积的方子阵的行列式有如下著名的 **Binet-Cauchy 公式**. 其证明见文献[2].

定理 3　设 $A\in \boldsymbol{M}_{n,p}(\boldsymbol{F}), B\in \boldsymbol{M}_{p,m}(\boldsymbol{F}), C=AB\in \boldsymbol{M}_{n,m}(\boldsymbol{F})$. 若 $1\leqslant r\leqslant \min\{n,m,p\}$，$\alpha\in Q_{r,n}$，$\beta\in Q_{r,m}$，则

$$\det C[\alpha \mid \beta] = \sum_{\omega\in Q_{r,p}} \det A[\alpha \mid \omega]\det B[\omega \mid \beta]. \tag{19}$$

假如取 $n=m=r=p$，则由(19)式推出如下有用的结果.

推论 4　设 $A,B\in \boldsymbol{M}_n(\boldsymbol{F})$，则

$$\det(AB) = \det A\cdot \det B. \tag{20}$$

假如 $A\in \boldsymbol{M}_n(\boldsymbol{F})$ 满足 $\det A\neq 0$，则称 A 为**非奇异矩阵**，否则称 A 为**奇异矩阵**. 若用 $\mathrm{adj}A$ 表示 $A=(a_{ij})\in \boldsymbol{M}_n(\boldsymbol{F})$ 的**伴随矩阵**，即 $\mathrm{adj}A=((A_{ij})_{i,j=1}^{n})^{\mathrm{T}}$，这里 A_{ij} 为 a_{ij} 的代数余子式，则由(15)式与(16)式得出

$$(\mathrm{adj}A)A = A(\mathrm{adj}A) = (\det A)I_n. \tag{21}$$

矩阵 $A\in \boldsymbol{M}_n(\boldsymbol{F})$ 称为**可逆的**，假若有 $B\in \boldsymbol{M}_n(\boldsymbol{F})$ 使得 $AB=BA=I_n$. 这时，B 叫作 A 的**逆**. 当 $\det A\neq 0$ 时，(21)式蕴涵 $A\in \boldsymbol{M}_n(\boldsymbol{F})$ 为可逆矩阵. 反之，若 $A\in \boldsymbol{M}_n(\boldsymbol{F})$ 可逆，则有 B 使得 $AB=I_n$，因而按推论 4，$\det(AB)=\det A\cdot\det B=1$，由此推出 $\det A\neq 0$. 因此，$A\in \boldsymbol{M}_n(\boldsymbol{F})$ 可逆当且仅当 A 为非奇异矩阵. 当 $A\in \boldsymbol{M}_n(\boldsymbol{F})$ 可逆时，其逆一定唯一，这是因为若有 B_1 与 B_2 使得 $B_1A=AB_1=AB_2=B_2A=I_n$，则 $A(B_1-B_2)=O$. 再用 B_1 左乘之得 $B_1A(B_1-B_2)=O$，这蕴涵 $B_1=B_2$. 今后将可逆矩阵 $A\in \boldsymbol{M}_n(\boldsymbol{F})$ 的唯一逆记为 A^{-1}. 按(21)式我们有

$$A^{-1} = \mathrm{adj}A/\det A, (A^{-1})^{\mathrm{T}} = (A^{\mathrm{T}})^{-1}, \tag{22}$$

$$\det(A^{-1}) = (\det A)^{-1}. \tag{23}$$

直接代入验证便知，若 $A,B\in \boldsymbol{M}_n(\boldsymbol{F})$ 可逆，则

$$(AB)^{-1} = B^{-1}A^{-1}, \tag{24}$$

并且，对任意整数 p，A^p 有意义，此外，当 p 与 q 为整数时，

$$A^pA^q = A^{p+q}, \quad (A^p)^q = A^{pq}. \tag{25}$$

一个值得注意的事实是，当 $A\in \boldsymbol{M}_n(\boldsymbol{F})$ 时，$AB=I_n$（即 A 有右逆 B）当且仅当 $BA=I_n$（A 有左逆 B），它又相当于：$B=A^{-1}$. 这是因为 $AB=I_n$ 蕴涵 $A\boldsymbol{x}=\boldsymbol{e}_1$（$\boldsymbol{e}_1=(1,0,\cdots,0)^{\mathrm{T}}\in\mathbb{R}^n$）以 B 的第一列向量 $B^{(1)}$ 为其解，但此方程组显然只有唯一解 $\boldsymbol{x}=A^{-1}\boldsymbol{e}_1=(A^{-1})^{(1)}$（因为 $AB=I_n$ 蕴涵 $\det A\neq 0$，即 A^{-1} 存在），于是，$B^{(1)}=(A^{-1})^{(1)}$. 按类似方法可证明 $B^{(i)}=(A^{-1})^{(i)}$，$i=2,\cdots,n$，即有 $B=A^{-1}$. 余下的命题均为显然的.

在乘法运算下，$\boldsymbol{M}_n(\boldsymbol{F})$中所有非奇异矩阵组成一个群——**一般线性群**，记为$GL(n,\boldsymbol{F})$. 事实上，若$A,B\in\boldsymbol{M}_n(\boldsymbol{F})$非奇异，则$B^{-1}A$恰是$A^{-1}B$的逆，因而$A^{-1}B$仍为非奇异的$n$阶矩阵.

现在介绍 **Schur 余量**概念. 设$\alpha\in Q_{r,n}$且$A\in\boldsymbol{M}_n(\boldsymbol{F})$. 若$A[\alpha]$非奇异，则有

$$\det A=\det A[\alpha]\cdot\det(A(\alpha)-A(\alpha\mid\alpha]A[\alpha]^{-1}A[\alpha\mid\alpha)). \tag{26}$$

事实上，不妨设A为如下对称分块矩阵：

$$A=\begin{bmatrix}A_{11}&A_{12}\\A_{21}&A_{22}\end{bmatrix},$$

其中$A_{11}\in\boldsymbol{M}_r(\boldsymbol{F})$可逆. 这时，(26)式的相应形式

$$\det A=\det A_{11}\cdot\det(A_{22}-A_{21}A_{11}^{-1}A_{12}) \tag{27}$$

可从下面关系式与推论 4 直接得出：

$$\begin{bmatrix}I_r&O\\A_{21}A_{11}^{-1}&I_{n-r}\end{bmatrix}\begin{bmatrix}A_{11}&O\\O&A_{22}-A_{21}A_{11}^{-1}A_{12}\end{bmatrix}\begin{bmatrix}I_r&A_{11}^{-1}A_{12}\\O&I_{n-r}\end{bmatrix}=A. \tag{28}$$

(26)式中$A(\alpha)-A(\alpha|\alpha]A[\alpha]^{-1}A[\alpha|\alpha)$叫作矩阵$A$内$A[\alpha]$的 **Schur 余量**(Schur complement)，记为$[A/A[\alpha]]$.

因此，上述分块矩阵A可逆，当且仅当A_{11}与$[A/A_{11}]$均可逆. 此时从(28)式推出：

$$\begin{aligned}A^{-1}&=\begin{bmatrix}I_r&-A_{11}^{-1}A_{12}\\O&I_{n-r}\end{bmatrix}\begin{bmatrix}A_{11}^{-1}&O\\O&(A_{22}-A_{21}A_{11}^{-1}A_{12})^{-1}\end{bmatrix}\begin{bmatrix}I_r&O\\-A_{21}A_{11}^{-1}&I_{n-r}\end{bmatrix}\\&=\begin{bmatrix}A_{11}^{-1}+A_{11}^{-1}A_{12}(A_{22}-A_{21}A_{11}^{-1}A_{12})^{-1}A_{21}A_{11}^{-1}&A_{11}^{-1}A_{12}(A_{21}A_{11}^{-1}A_{12}-A_{22})^{-1}\\(A_{21}A_{11}^{-1}A_{12}-A_{22})^{-1}A_{21}A_{11}^{-1}&(A_{22}-A_{21}A_{11}^{-1}A_{12})^{-1}\end{bmatrix}.\end{aligned} \tag{29}$$

定理 5(矩阵的LU分解)　设$A=(a_{ij})\in\boldsymbol{M}_n(\boldsymbol{F})$的头$n-1$个前主子式不为零，即有$\det A[(1)]\neq0,\det A[(1,2)]\neq0,\cdots,\det A[(1,\cdots,n-1)]\neq0$. 则存在唯一下三角形矩阵$L=(l_{ij})$，其所有对角元素$l_{ii}=1$，与唯一上三角形矩阵$U=(u_{ij})$，使得$A=LU$与$\det A=\det U=u_{11}u_{22}\cdots u_{nn}$，这里$u_{11}=a_{11}$，$u_{kk}=\det A[(1,\cdots,k)]/\det A[(1,\cdots,k-1)]$，$k=2,\cdots,n$.

证明　对A的阶数n应用归纳法. 如$n=1$，则结论显然成立，假定本结论对$n-1$阶矩阵成立，考虑n阶矩阵情形. 将A写成对称分块矩阵形式

$$A=\begin{bmatrix}A_{n-1}&\boldsymbol{a}\\\boldsymbol{b}^{\mathrm{T}}&a_{nn}\end{bmatrix},$$

这里，$\boldsymbol{a},\boldsymbol{b}\in\boldsymbol{F}^{n-1}$. 按归纳假定，$A_{n-1}$有唯一三角分解：$A_{n-1}=\widetilde{L}\widetilde{U}$，其中$\widetilde{L}$为对角元素皆等于 1 的$n-1$阶下三角形矩阵，$\widetilde{U}$为$n-1$阶上三角形矩阵. 由于$\det A_{n-1}=\det A[(1,\cdots,n-1)]\neq0$，故$\widetilde{L}$与$\widetilde{U}$都是非奇异的. 现令

$$L=\begin{bmatrix}\widetilde{L} & \mathbf{0}\\ \boldsymbol{b}^{\mathrm{T}}\widetilde{U}^{-1} & 1\end{bmatrix}\quad 与 \quad U=\begin{bmatrix}\widetilde{U} & \widetilde{L}^{-1}\boldsymbol{a}\\ \boldsymbol{o}^{\mathrm{T}} & a_{m}-\boldsymbol{b}^{\mathrm{T}}A_{n-1}^{-1}\boldsymbol{a}\end{bmatrix}. \tag{30}$$

按(27)式,我们有

$$a_m-\boldsymbol{b}^{\mathrm{T}}A_{n-1}^{-1}\boldsymbol{a}=\det(a_m-\boldsymbol{b}^{\mathrm{T}}A_{n-1}^{-1}\boldsymbol{a})=\det A/\det A_{n-1}.$$

容易验证,(30)式中的 L 与 U 便是满足条件 $A=LU$ 的下三角形与上三角形矩阵. 因为 $\det L=1$,根据推论 4 有 $\det A=\det U$. 现在证明这样三角分解 $A=LU$ 的唯一性. 假定

$$\hat{L}=\begin{bmatrix}L_{11} & \mathbf{0}\\ \boldsymbol{d}^{\mathrm{T}} & 1\end{bmatrix}\quad 与 \quad \hat{U}=\begin{bmatrix}U_{11} & \boldsymbol{h}\\ \mathbf{0}^{\mathrm{T}} & u_m\end{bmatrix}$$

为满足 $A=\hat{L}\hat{U}$ 的上三角形与下三角形矩阵,$L_{11}\in\boldsymbol{M}_{n-1}(\boldsymbol{F})$的对角元素皆为 1. 按分块矩阵乘法可得

$$A_{n-1}=L_{11}U_{11},\quad \boldsymbol{d}^{\mathrm{T}}U_{11}=\boldsymbol{b}^{\mathrm{T}},\quad L_{11}\boldsymbol{h}=\boldsymbol{a},\quad \boldsymbol{d}^{\mathrm{T}}\boldsymbol{h}+u_m=a_m. \tag{31}$$

根据归纳假定,$L_{11}=\widetilde{L}$,$U_{11}=\widetilde{U}$;根据 $\widetilde{U}$ 与 $\widetilde{L}$ 的非奇异性,由(31)式中第二个与第三个方程解出 $\boldsymbol{d}^{\mathrm{T}}=\boldsymbol{b}^{\mathrm{T}}\widetilde{U}^{-1}$ 与 $\boldsymbol{h}=\widetilde{L}^{-1}\boldsymbol{a}$,因而从(31)式的第四个方程得出 $u_m=a_m-\boldsymbol{b}^{\mathrm{T}}A_{n-1}^{-1}\boldsymbol{a}$. 因此,$\hat{L}=L$ 与 $\hat{U}=U$. □

我们指出,定理 5 的逆命题也成立,即若 $A\in\boldsymbol{M}_n(\boldsymbol{F})$有如定理 5 所述的三角分解:$A=LU$,这里 U 的前$n-1$个对角元素均不为零,则 A 的头 $n-1$ 个前主子式不为零.

定理 5 还可以推广到分块矩阵的情形,详请见习题 10.

有了行列式的概念,可以讨论矩阵 $A\in\boldsymbol{M}_{m,n}(\boldsymbol{F})$ 的**秩**,记为 $\operatorname{rank}A$. $A\in\boldsymbol{M}_{m,n}(\boldsymbol{F})$的秩定义为 A 的最大非奇异子阵的阶数. 因此,除了 $A=O_{m,n}$有 $\operatorname{rank}A=0$ 外,非零矩阵 A 都有 $\operatorname{rank}A\geqslant 1$. 一般地,$\operatorname{rank}A\leqslant\min\{m,n\}$,$\forall A\in\boldsymbol{M}_{m,n}(\boldsymbol{F})$. A 的秩也等于 A 的所有行向量(或列向量)中线性无关向量的最大个数,也就是说,$\operatorname{rank}A$ 等于 A 的**行空间**(或**列空间**)的维数. 这里,当 $A\in\boldsymbol{M}_{m,n}(\boldsymbol{F})$时,$A$ 的列空间即为 A 的**值域**,记为 $\operatorname{Im}A$,它由下式确定:

$$\operatorname{Im}A=\{\boldsymbol{y}\in\boldsymbol{F}^m:\boldsymbol{y}=A\boldsymbol{x},\forall\boldsymbol{x}\in\boldsymbol{F}^n\}; \tag{32}$$

A 的行空间等于 $\operatorname{Im}A^{\mathrm{T}}$. 因此,$\operatorname{Im}A$ 为 $\boldsymbol{F}^m$ 的线性子空间,并且

$$\operatorname{rank}A=\dim(\operatorname{Im}A)=\dim(\operatorname{Im}A^{\mathrm{T}}). \tag{33}$$

若 $A\in\boldsymbol{M}_{m,n}(\boldsymbol{F})$满足 $\operatorname{rank}A=m(\leqslant n)$,则称 A 为**行满秩的**;若它满足 $\operatorname{rank}A=n(\leqslant m)$,则称 A 为**列满秩的**. 因此,当且仅当 $A\in\boldsymbol{M}_n(\boldsymbol{F})$非奇异时,$A$ 是满秩的,即 $\operatorname{rank}A=n$. 矩阵秩的最重要的性质罗列如下.

(1) (**Frobenius 不等式**)若 $A\in\boldsymbol{M}_{m,k}(\boldsymbol{F})$,$B\in\boldsymbol{M}_{k,p}(\boldsymbol{F})$,$C\in\boldsymbol{M}_{p,n}(\boldsymbol{F})$,则

$$\operatorname{rank}(AB)+\operatorname{rank}(BC)\leqslant\operatorname{rank}B+\operatorname{rank}(ABC). \tag{34}$$

特殊地,先取 $C=O_{p,n}$,再取 $A=O_{m,k}$,由上述不等式可得

$$\mathrm{rank}(AB) \leqslant \min\{\mathrm{rank}A, \mathrm{rank}B\}. \tag{35}$$

(2) 若 $A \in \boldsymbol{M}_{m,n}(\boldsymbol{F})$，则

$$\mathrm{rank}A = \mathrm{rank}(AA^*) = \mathrm{rank}(A^*A). \tag{36}$$

(3) 若 $P \in \boldsymbol{M}_m(\boldsymbol{F})$ 与 $Q \in \boldsymbol{M}_n(\boldsymbol{F})$ 非奇异，且 $A \in \boldsymbol{M}_{m,n}(\boldsymbol{F})$，则

$$\mathrm{rank}(PA) = \mathrm{rank}A = \mathrm{rank}(AQ) = \mathrm{rank}(PAQ). \tag{37}$$

(4) $\mathrm{rank}A = \mathrm{rank}A^{\mathrm{T}} = \mathrm{rank}A^*$.

(5) 若 $A, B \in \boldsymbol{M}_{m,n}(\boldsymbol{F})$，则

$$\mathrm{rank}(A+B) \leqslant \mathrm{rank}A + \mathrm{rank}B. \tag{38}$$

(6) $\mathrm{rank}\ \ \mathrm{diag}[A_{11}, A_{22}, \cdots, A_{kk}] = \sum_{j=1}^{k} \mathrm{rank}A_{jj}$.

(7) 若 $A \in \boldsymbol{M}_{m,n}(\boldsymbol{F})$ 有秩 k，则有 $X \in \boldsymbol{M}_{m,k}(\boldsymbol{F})$，$Y \in \boldsymbol{M}_{k,n}(\boldsymbol{F})$ 与非奇异 $B \in \boldsymbol{M}_k(\boldsymbol{F})$，使得 $A = XBY$. 特殊地，若 $\mathrm{rank}A = 1$，则有 $\boldsymbol{x} \in \boldsymbol{F}^m$ 与 $\boldsymbol{y} \in \boldsymbol{F}^n$ 使得 $A = \boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}$.

假如已经知道非奇异矩阵 $A \in \boldsymbol{M}_n(\boldsymbol{F})$ 的逆 A^{-1}，考虑

$$B = A + XCY, \tag{39}$$

这里，$X \in \boldsymbol{M}_{n,r}(\boldsymbol{F})$，$Y \in \boldsymbol{M}_{r,n}(\boldsymbol{F})$ 且 $C \in \boldsymbol{M}_r(\boldsymbol{F})$ 非奇异. 若 B 非奇异，则应用(28)式便知

$$B^{-1} = A^{-1} - A^{-1}X(C^{-1} + YA^{-1}X)^{-1}YA^{-1}. \tag{40}$$

(见习题 15)特殊地，若 C 的秩 r 比 n 小得多，则 C^{-1} 与 $(C^{-1}+YA^{-1}X)^{-1}$ 往往较为容易求得，例如当 $r=1$ 与 $C=(1)$ 时，我们有

$$B^{-1} = A^{-1} - \frac{1}{1 + YA^{-1}X}A^{-1}XYA^{-1}, \tag{41}$$

注意这时 $XY = B - A$. 因此，若 $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{x} \neq -1$，且

$$B = I_n - \boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}, \boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{F}^n, \tag{42}$$

则有

$$B^{-1} = I_n + \frac{1}{1 + \boldsymbol{y}^{\mathrm{T}}\boldsymbol{x}}\boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}. \tag{43}$$

应用公式(40)，当 A_{22} 也是可逆时，(29)式可改写为

$$A^{-1} = \begin{bmatrix} (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} & A_{11}^{-1}A_{12}(A_{21}A_{11}^{-1}A_{12} - A_{22})^{-1} \\ (A_{21}A_{11}^{-1}A_{12} - A_{22})^{-1}A_{21}A_{11}^{-1} & (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1} \end{bmatrix}. \tag{29'}$$

与矩阵 $A \in \boldsymbol{M}_{m,n}(\boldsymbol{F})$ 的值域 $\mathrm{Im}A$ 紧密相关的是 A 的**核**(Kernel)或**零空间**(null space)，记为 $\mathrm{Ker}A$，它是 $\boldsymbol{F}^n$ 的如下线性子空间

$$\mathrm{Ker}A = \{\boldsymbol{x} \in \boldsymbol{F}^n : A\boldsymbol{x} = \boldsymbol{0}\}. \tag{44}$$

因为，$\mathrm{Ker}A$ 即为方程组 $A\boldsymbol{x} = \boldsymbol{0}$ 的解空间. 假如 AB 有定义，按(44)式我们有 $\mathrm{Ker}(AB) \supset \mathrm{Ker}B$，且当 A^{-1} 存在时，$\mathrm{Ker}(AB) = \mathrm{Ker}B$. 当 $A \in \boldsymbol{M}_n(\boldsymbol{F})$ 时 $A(\mathrm{Ker}A) \subset \mathrm{Ker}A$，即 $\mathrm{Ker}A$ 为 A 的不变子空间. 顺便地说，对于矩阵的值域也有类

似的结果:假如 AB 有定义,则 $\mathrm{Im}(AB)\subset\mathrm{Im}A$,且当 B^{-1} 存在时,$\mathrm{Im}(AB)=\mathrm{Im}A$. 当 $A\in\boldsymbol{M}_n(\boldsymbol{F})$ 时,$A(\mathrm{Im}A)\subset\mathrm{Im}A$,即 $\mathrm{Im}A$ 为 A 的不变子空间.

$\mathrm{Ker}A$ 的维数称为 A 的**零度**(nullity),记为 $\mathrm{nul}A$. 显然,非奇异矩阵的零度为 0,奇异矩阵的零度大于 0.

定理 6 若 $A\in\boldsymbol{M}_{m,n}(\boldsymbol{F})$,则

$$\mathrm{rank}A+\mathrm{nul}A=n. \tag{45}$$

证明 设 $\mathrm{rank}A=r$,且不妨设 A 可写成如下的分块矩阵形式(因为交换 A 的行与列不改变 A 的秩):

$$A=\begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix},$$

式中 $A_{11}\in\boldsymbol{M}_r(\boldsymbol{F})$ 非奇异. 由于 $\mathrm{rank}A=r$,故存在 $P\in\boldsymbol{M}_m(\boldsymbol{F})$ 与 $Q\in\boldsymbol{M}_n(\boldsymbol{F})$,它们为一些初等变换矩阵的乘积,因而是非奇异的,使得

$$P^{-1}AQ^{-1}=\begin{bmatrix}I_r & O\\ O & O\end{bmatrix}\equiv\widetilde{A}\in\boldsymbol{M}_{m,n}(\boldsymbol{F}). \tag{46}$$

此时,$\mathrm{rank}\widetilde{A}=\mathrm{rank}A=r$ 且 $\mathrm{nul}\widetilde{A}=n-r$,因为 $\widetilde{A}\boldsymbol{x}=\boldsymbol{0}$ 恰有 $n-r$ 个线性无关的解向量. 余下只要证明 $\mathrm{nul}\widetilde{A}=\mathrm{nul}A$ 即可. 若 $\boldsymbol{x}\in\mathrm{Ker}\widetilde{A}$,则由于 $AQ^{-1}=P\widetilde{A}$,故我们有 $Q^{-1}\boldsymbol{x}\in\mathrm{ker}A$. 反之,$Q^{-1}\boldsymbol{x}\in\mathrm{Ker}A$ 蕴涵 $\boldsymbol{x}\in\mathrm{Ker}\widetilde{A}$,因为 P 为非奇异的. 于是,$\mathrm{Ker}A=Q^{-1}(\mathrm{Ker}\widetilde{A})$. 但 $Q^{-1}(\mathrm{Ker}\widetilde{A})$ 与 $\mathrm{Ker}\widetilde{A}$ 的维数相等,因而 $\mathrm{nul}A=\mathrm{nul}\widetilde{A}$. □

若 $A=(a_{ij})\in\boldsymbol{M}_n(\mathbb{C})$,则 n 次首一的多项式

$$\Delta(\lambda)\equiv\det(\lambda I_n-A) \tag{47}$$

称为 A 的**特征多项式**. 当 $1\leqslant r\leqslant n$ 时,记

$$E_r(A)\equiv\sum_{\omega\in Q_{r,n}}\det A[\omega], \tag{48}$$

则有 $E_1(A)=\sum_{i=1}^{n}a_{ii}$,它为 A 的**迹**(trace),记为 $\mathrm{tr}A$;$E_n(A)=\det A$. 应用(48)式,我们有(约定 $E_0(A)=1$)

$$\Delta(\lambda)=\det(\lambda I_n-A)=\sum_{t=0}^{n}\lambda^{n-t}(-1)^tE_t(A). \tag{49}$$

例如,当 $n=2$ 时,$\Delta(\lambda)=\det(\lambda I_2-A)=\lambda^2-(\mathrm{tr}A)\lambda+\det A$.

A 的特征多项式 $\Delta(\lambda)=\det(\lambda I_n-A)$ 的 n 个零点(计入重数)叫作 A 的**特征值**. 因此,若 A 的特征值为 $\lambda_1,\cdots,\lambda_n$,则

$$\Delta(\lambda)=\det(\lambda I_n-A)=\prod_{j=1}^{n}(\lambda-\lambda_j). \tag{50}$$

由此推出

$$\mathrm{tr}A=\sum_{i=1}^{n}a_{ii}=\sum_{i=1}^{n}\lambda_i,\quad \det A=\prod_{i=1}^{n}\lambda_i. \tag{51}$$

特征值的某些基本性质罗列如下.

(1) 若 $A=\begin{bmatrix} B & C \\ O & D \end{bmatrix}\in \boldsymbol{M}_n(\mathbb{C})$，$B\in \boldsymbol{M}_p(\mathbb{C})$，$D\in \boldsymbol{M}_q(\mathbb{C})$，$p+q=n$，则 A 的特征值为 B 的 p 个特征值与 D 的 q 个特征值. 特殊地，上三角形或下三角形复矩阵的特征值等于它的主对角线上的元素.

(2) 若 $A\in \boldsymbol{M}_n(\mathbb{C})$ 有特征值 λ_i，$i=1,\cdots,n$，则对任意 $\alpha\in\mathbb{C}$，αA 有特征值 $\alpha\lambda_i$，$i=1,\cdots,n$；对任意正整数 p，A^p 有特征值 λ_i^p，$i=1,\cdots,n$. 因此，若 $a_0,\cdots,a_m\in\mathbb{C}$ 且 $B=a_0I_n+a_1A+\cdots+a_mA^m$，则 B 的特征值为

$$a_0+a_1\lambda_i+a_2\lambda_i^2+\cdots+a_m\lambda_i^m,\quad i=1,\cdots,n.$$

(3) 若 $A\in \boldsymbol{M}_n(\mathbb{C})$，则 A 非奇异当且仅当 $\lambda_i\neq 0$，$i=1,\cdots,n$. 这时，若 $p\neq 0$ 为任意整数，则 A^p 的特征值为 λ_i^p，$i=1,\cdots,n$.

(4) 若 $A\in \boldsymbol{M}_n(\mathbb{C})$，则对任意非奇异 $S\in \boldsymbol{M}_n(\mathbb{C})$，$S^{-1}AS$ 与 A 有相同的特征值. 由此推出，若 $A\in \boldsymbol{M}_{m,n}(\mathbb{C})$ 与 $B\in \boldsymbol{M}_{n,m}$，则 $AB\in \boldsymbol{M}_m(\mathbb{C})$ 与 $BA\in \boldsymbol{M}_n(\mathbb{C})$ 有相同的非零特征值这个事实得自下列公式：

$$\begin{bmatrix} I_m & -A \\ O & I_n \end{bmatrix}\begin{bmatrix} AB & O \\ B & O \end{bmatrix}\begin{bmatrix} I_m & A \\ O & I_n \end{bmatrix}=\begin{bmatrix} O & O \\ B & BA \end{bmatrix}. \tag{52}$$

(5) 若 $A_1,\cdots,A_k\in \boldsymbol{M}_n(\mathbb{C})$ 满足 $A_iA_j=A_jA_i$，$\forall i,j$，且 $p(\lambda_1,\cdots,\lambda_k)$ 为具有复系数的关于 $\lambda_1,\cdots,\lambda_k$ 的多项式，则 $B=p(A_1,\cdots,A_k)$ 的特征值等于 $\beta_j=p(\lambda_{1j},\lambda_{2j},\cdots,\lambda_{kj})$，$j=1,\cdots,n$，这里 $\lambda_{ij}(1\leqslant j\leqslant n)$ 为 A_i 的特征值，它们按某个固定次序排列. 特殊地，若 $A_1A_2=A_2A_1$ 则 A_1A_2 的特征值为 $\lambda_{11}\lambda_{21},\lambda_{12}\lambda_{22},\cdots,\lambda_{1n}\lambda_{2n}$，其中，$\lambda_{11},\cdots,\lambda_{1n}$ 与 $\lambda_{21},\cdots,\lambda_{2n}$ 为 A_1 与 A_2 按某种次序排列的特征值.

(6) 若 $A\in \boldsymbol{M}_n(\mathbb{C})$ 为幂等矩阵，则 A 的特征值等于 0 或 1；若 $A\in \boldsymbol{M}_n(\mathbb{C})$ 为幂零矩阵，则 A 的特征值均为零.

若 $A\in \boldsymbol{M}_n(\mathbb{C})$，且 r 为 A 的特征值，则 $\det(rI_n-A)=0$，因而 $A-rI_n$ 为奇异矩阵. 这时，$\mathrm{nul}(A-rI_n)$ 称为特征值 r 的**几何重数**，而 $\mathrm{Ker}(A-rI_n)$ 中任意非零向量 $\boldsymbol{x}$ 称为对应于 r 的 A 的**特征向量**. 特征值 r 的**代数重数**恰为 A 的特征多项式 $\det(\lambda I_n-A)$ 中零点 r 的重数. 若 $A\in \boldsymbol{M}_n(\mathbb{C})$ 有不同特征值 $\lambda_1,\lambda_2,\cdots,\lambda_k$ 且 $\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_k$ 为对应的特征向量：$A\boldsymbol{u}_j=\lambda_j\boldsymbol{u}_j$，$j=1,\cdots,k$，则 $\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_k$ 线性无关. 若 $A\in \boldsymbol{M}_n(\mathbb{C})$ 有 n 个线性无关特征向量 $\boldsymbol{u}_1,\cdots,\boldsymbol{u}_n$，$A\boldsymbol{u}_j=\lambda_j\boldsymbol{u}_j$，$j=1,\cdots,n$，则有

$$A[\boldsymbol{u}_1\ \ \cdots\ \ \boldsymbol{u}_n]=[\boldsymbol{u}_1\ \ \cdots\ \ \boldsymbol{u}_n]\mathrm{diag}[\lambda_1,\cdots,\lambda_n]. \tag{53}$$

由于 $U\equiv[\boldsymbol{u}_1\ \ \cdots\ \ \boldsymbol{u}_n]\in \boldsymbol{M}_n(\mathbb{C})$ 非奇异，上式表明存在 $U\in \boldsymbol{M}_n(\boldsymbol{F})$ 使得 $U^{-1}AU=D=\mathrm{diag}[\lambda_1,\cdots,\lambda_n]$. 反之，若有 $U\in \boldsymbol{M}_n(\boldsymbol{F})$ 使得 $U^{-1}AU=D=\mathrm{diag}[\lambda_1,\cdots,\lambda_n]$，则由(53)式看出，$\lambda_j$ 与 $\boldsymbol{u}_j$（U 的第 j 个列向量）分别为 A 的特征值与对应的特征向量，$j=1,\cdots,n$. 今后将这样的有 n 个线性无关特征向量的矩阵 $A\in \boldsymbol{M}_n(\mathbb{C})$ 称为**简单矩阵或可对角化矩阵**(diagonalization-matrix). 由前面知道，若 $A\in \boldsymbol{M}_n(\mathbb{C})$ 有 n

个不同特征值,则 A 为简单矩阵. 但反之不真,例如 $A=I_n$.

定理 7(简单矩阵的谱定理) 设 $A\in \boldsymbol{M}_n(\mathbb{C})$ 为简单矩阵,有特征值 $\lambda_1,\cdots,\lambda_n$ 与对应的特征向量 $\boldsymbol{u}_1,\cdots,\boldsymbol{u}_n$. 则存在 n 阶矩阵 $G_j, j=1,\cdots,n$,使得 $\mathrm{rank}G_j=1$, $G_kG_j=\delta_{kj}G_k$, $\forall j,k$,并且,A 有如下的谱形式:

$$A=\sum_{j=1}^{n}\lambda_jG_j. \tag{54}$$

证明 令 $P=[\boldsymbol{u}_1\quad \boldsymbol{u}_2\quad \cdots\quad \boldsymbol{u}_n]$ 与 $Q=(P^{-1})^{\mathrm{T}}=[\boldsymbol{v}_1\quad \boldsymbol{v}_2\quad \cdots\quad \boldsymbol{v}_n]$. 则由 $Q^{\mathrm{T}}P=I_n$ 得出 $\boldsymbol{v}_j^{\mathrm{T}}\boldsymbol{u}_k=\delta_{jk}$, $\forall k,j\in \mathcal{N}=\{1,\cdots,n\}$. 从 $A=PDP^{-1}$, $D=\mathrm{diag}[\lambda_1,\cdots,\lambda_n]$推出 $A^{\mathrm{T}}Q=QD$,这表明 Q 的列向量 $\boldsymbol{v}_j$ 为对应于 λ_j 的 A^{T} 的特征向量:$A^{\mathrm{T}}\boldsymbol{v}_j=\lambda_j\boldsymbol{v}_j$, $j=1,\cdots,n$. 令 $G_j=\boldsymbol{u}_j\boldsymbol{v}_j^{\mathrm{T}}$, $j=1,\cdots,n$. 则从

$$A=PDQ^{\mathrm{T}}=[\boldsymbol{u}_1\quad \boldsymbol{u}_2\quad \cdots\quad \boldsymbol{u}_n]\begin{bmatrix}\lambda_1\boldsymbol{v}_1^{\mathrm{T}}\\ \lambda_2\boldsymbol{v}_2^{\mathrm{T}}\\ \vdots\\ \lambda_n\boldsymbol{v}_n^{\mathrm{T}}\end{bmatrix}=\sum_{j=1}^{n}\lambda_j\boldsymbol{u}_j\boldsymbol{v}_j^{\mathrm{T}}$$

直接得到(54)式,因为这时 $\mathrm{rank}G_j=1$, $j=1,\cdots,n$,且 $G_kG_j=\boldsymbol{u}_k\boldsymbol{v}_k^{\mathrm{T}}\boldsymbol{u}_j\boldsymbol{v}_j^{\mathrm{T}}=\delta_{kj}G_k$, $\forall k,j\in\mathcal{N}$. □

我们指出,在定理 7 证明中,$A^{\mathrm{T}}\boldsymbol{v}_j=\lambda_j\boldsymbol{v}_j$ 蕴涵 $\boldsymbol{v}_j^{\mathrm{T}}A=\lambda_j\boldsymbol{v}_j^{\mathrm{T}}$, $j=1,\cdots,n$,因而 $\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n$ 可叫作 A 的**左特征向量**. 这样一来,定理 7 表示 A 可由它的**谱性质**,即由 A 的特征值、特征向量与左特征向量确定. 在第三章 3.1.2 小节中我们将讨论简单矩阵的谱定理的推广与应用.

习题 1.1.1

1. 试证:若 $A\in \boldsymbol{M}_{m,k}(\boldsymbol{F})$, $B\in \boldsymbol{M}_{k,n}(\boldsymbol{F})$, $C=AB$. 则 $C_{(i)}=A_{(i)}B(i=1,\cdots,m)$, $C^{(j)}=AB^{(j)}(j=1,\cdots,n)$.

2. 试证:若 $A=(a_{ij})\in \boldsymbol{M}_n(\boldsymbol{F})$ 中 $a_{i,i+1}\neq 0$, $i=1,\cdots,n-1$,其余元素均为零,则 A 为幂零矩阵,且 $A^n=O$.

3. 试证:若 $A=(a_{ij})\in \boldsymbol{M}_n(\mathbb{C})$,则 $\mathrm{tr}(AA^*)=\mathrm{tr}(A^*A)=\sum_{i,j=1}^{n}|a_{ij}|^2$;若 $A,B\in \boldsymbol{M}_n(\mathbb{C})$,则 $\mathrm{tr}(\alpha A+\beta B)=\alpha\mathrm{tr}A+\beta\mathrm{tr}B$, $\forall \alpha,\beta\in\mathbb{C}$ 若 $A\in \boldsymbol{M}_{m,n}(\mathbb{C})$, $B\in \boldsymbol{M}_{n,m}(\mathbb{C})$,则有 $\mathrm{tr}(AB)=\mathrm{tr}(BA)$;若 $A,B\in \boldsymbol{M}_n(\mathbb{C})$且 A 为幂等的,则 $\mathrm{tr}(AB)=\mathrm{tr}(ABA)$;若 $n\geqslant 2$,则不存在 $A,B\in \boldsymbol{M}_n(\mathbb{C})$使得

$$AB-BA=I_n.$$

4. 试用数学归纳法证明:若 $A,B\in \boldsymbol{M}_n(\boldsymbol{F})$,则

$$A^r-B^r=\sum_{j=0}^{r-1}B^j(A-B)A^{r-1-j},\quad r=1,2,\cdots.$$

5. 设 $A\in \boldsymbol{M}_n(\boldsymbol{F})$有分块矩阵形式

$$A=\begin{bmatrix}A_{11} & A_{12}\\ A_{21} & O\end{bmatrix},$$

这里,A_{12}与A_{21}分别为p阶与$n-p$阶矩阵,$p<n$. 试证:$\det A=(-1)^{(n+1)p}\det A_{12}.\det A_{21}$.

6. 试证:若$A\in \boldsymbol{M}_{m,n}(\boldsymbol{F})$的所有$k$阶子式均为零,$1\leqslant k<\min\{m,n\}$,则$\operatorname{rank}A\leqslant k-1$.

7. 应用 Binet-Cauchy 公式证明:

(1) 若$A\in \boldsymbol{M}_{k,n}(\boldsymbol{F})$,$k\leqslant n$,则$\det(AA^*)\geqslant 0$.

(2) **Cauchy 恒等式**:

$$\Big(\sum_{i=1}^{n}a_ic_i\Big)\Big(\sum_{i=1}^{n}b_id_i\Big)-\Big(\sum_{i=1}^{n}a_id_i\Big)\Big(\sum_{i=1}^{n}b_ic_i\Big)=\sum_{1\leqslant j<k\leqslant n}(a_jb_k-a_kb_j)(c_jd_k-c_kd_j),$$

其中,诸$a_i,b_i,c_i,d_i\in \boldsymbol{F}$.

(3) **Cauchy-Schwarz 不等式**:

$$\Big(\sum_{i=1}^{n}a_ib_i\Big)^2\leqslant\Big(\sum_{i=1}^{n}a_i^2\Big)\Big(\sum_{i=1}^{n}b_i^2\Big),$$

其中,诸$a_i,b_i\in\mathbb{R}$.

8. 设$A\in \boldsymbol{M}_n(\boldsymbol{F})$. $\boldsymbol{x},\boldsymbol{y}\in \boldsymbol{F}^n$. 试证:对任意$\alpha\in\boldsymbol{F}$,$\xi\in\boldsymbol{F}^n$,

$$\det\begin{bmatrix}A & \boldsymbol{x}\\ \boldsymbol{y}^{\mathrm{T}} & \alpha\end{bmatrix}=\alpha\det A-y^{\mathrm{T}}(\operatorname{adj}A)\boldsymbol{x},$$

$$\det\begin{bmatrix}A & A\xi\\ \xi^{\mathrm{T}}A & \alpha\end{bmatrix}=(\det A)(\alpha-\xi^{\mathrm{T}}A\xi).$$

9. 设$A=(a_{ij})\in \boldsymbol{M}_n(\mathbb{C})$可表示为$A=[A_1\quad A_2]$,其中$A_1,A_2$分别为$n\times k,n\times(n-k)$矩阵,$1\leqslant k\leqslant n-1$. 试证:

(1) $|\det A|^2\leqslant(\det A_1^*A_1)(\det A_2^*A_2)$.

(2) $|\det A|^2\leqslant\prod_{j=1}^{n}\Big(\sum_{i=1}^{n}|a_{ij}|^2\Big)$ (**Hadamard 不等式**).

(3) 当$|a_{ij}|\leqslant M(1\leqslant i,j\leqslant n)$时,$|\det A|\leqslant M^nn^{n/2}$.

*10. 设$A\in \boldsymbol{M}_n(\boldsymbol{F})$有对称分块矩阵形式$A=[A_{ij}]_{i,j=1}^{k}$. 试证:若$A_1=A_{11}$,$A_2=[A_{ij}]_{i,j=1}^{2},\cdots,A_{k-1}=[A_{ij}]_{i,j=1}^{k-1}$皆为可逆的,则存在分块下三角形矩阵$L$(其$k$个对角块皆为单位矩阵)与分块上三角形矩阵$U$使得$A=LU$. 并且,这样的$L$与$U$是唯一的,$U$的对角块$U_{jj}(1\leqslant j\leqslant k)$为

$U_{11}=A_{11}$,$U_{pp}=[[A_{ij}]_{i,j=1}^{p}/[A_{ij}]_{i,j=1}^{p-1}]$,$p=2,\cdots,k$,　其中,$U_{11},\cdots,U_{k-1,k-1}$非奇异.

11. 设A为对称分块矩阵$A=[A_{ij}]_{i,j=1}^{3}$,其中A_{11}与$\widetilde{A}=[A_{ij}]_{i,j=1}^{2}$均为非奇异的. 试证:

$$[A/\widetilde{A}]=[[A/A_{11}]/[\widetilde{A}/A_{11}]].$$

12. 设$A,B\in \boldsymbol{M}_{m,n}(\boldsymbol{F})$. 试证:

$$\operatorname{rank}(A+B)\leqslant\operatorname{rank}A+\operatorname{rank}B.$$

并且,上式等式成立当且仅当有非奇异矩阵$P\in \boldsymbol{M}_m(\boldsymbol{F})$,$Q\in \boldsymbol{M}_n(\boldsymbol{F})$,以及矩阵$A_0,B_0$使得

$$A=P\begin{bmatrix}A_0 & O\\ O & O\end{bmatrix}Q,\quad B=P\begin{bmatrix}O & O\\ O & B_0\end{bmatrix}Q,$$

这里,等式右端中间两个矩阵有一致的分块形式,或等价地,$\operatorname{Im}(A+B)$为$\operatorname{Im}A$与$\operatorname{Im}B$的直和.

13. 证明:n阶友阵

$$\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & & 0 & 1 \\ -c_0 & -c_1 & -c_2 & \cdots & -c_{n-1} \end{bmatrix}$$

的特征多项式为 $c(\lambda)=\lambda^n+\sum\limits_{i=0}^{n-1}c_i\lambda^i$. 并且,它的行列式等于 $(-1)^n c_0$.

14. 设简单矩阵 $A\in \boldsymbol{M}_n(\boldsymbol{F})$ 有表达式(54). 试证:若 $p(\lambda)$ 为复系数多项式,则 $p(A)$ 为简单矩阵,且 $p(A)=\sum\limits_{j=1}^{n}p(\lambda_j)G_j$;若 $c(\lambda)$ 为 A 的特征多项式,则 $c(A)=O$.

15. 设 $B=A+XCY$,其中 $A\in \boldsymbol{M}_n(\boldsymbol{F})$ 可逆. $X\in \boldsymbol{M}_{n,r}(\boldsymbol{F})$, $Y\in \boldsymbol{M}_{r,n}(\boldsymbol{F})$,且 $C\in \boldsymbol{M}_r(\boldsymbol{F})$ 可逆. 证明:若 B 可逆,则 $C^{-1}+YA^{-1}X$ 可逆,且 $B^{-1}=A^{-1}-A^{-1}X(C^{-1}+YA^{-1}X)^{-1}YA^{-1}$.

1.1.2　线性变换与矩阵表示,相似性与 Jordan 正规形式

设 $\mathcal{S}_1$ 与 $\mathcal{S}_2$ 为域 $\boldsymbol{F}$ 上两个有限维线性空间, $\dim\mathcal{S}_1=n$ 与 $\dim\mathcal{S}_2=m$. 若 $T:\mathcal{S}_1\to\mathcal{S}_2$ 为满足条件

$$T(\alpha\boldsymbol{x}+\beta\boldsymbol{y})=\alpha T(\boldsymbol{x})+\beta T(\boldsymbol{y}),\quad \forall\,\boldsymbol{x},\boldsymbol{y}\in\mathcal{S}_1,\alpha,\beta\in\boldsymbol{F}$$

的映射,则称 T 为自 $\mathcal{S}_1$ 到 $\mathcal{S}_2$ 内的一个**线性变换**. 假如 $\mathcal{S}_1=\mathcal{S}_2$,则 T 简称为 $\mathcal{S}_1$ 上的一个线性变换. 用 $\mathcal{L}(\mathcal{S}_1,\mathcal{S}_2)$ 表示自 $\mathcal{S}_1$ 到 $\mathcal{S}_2$ 内所有线性变换的集合,并记 $\mathcal{L}(\mathcal{S})\equiv\mathcal{L}(\mathcal{S},\mathcal{S})$. 若在 $\mathcal{L}(\mathcal{S}_1,\mathcal{S}_2)$ 内引入加法运算:

$$(T_1+T_2)(\boldsymbol{x})\equiv T_1(\boldsymbol{x})+T_2(\boldsymbol{x}),\qquad \forall\,T_1,T_2\in\mathcal{L}(\mathcal{S}_1,\mathcal{S}_2),\boldsymbol{x}\in\mathcal{S}_1$$

与数乘运算:

$$(\alpha T)(\boldsymbol{x})\equiv\alpha T(\boldsymbol{x}),\qquad \forall\,T\in\mathcal{L}(\mathcal{S}_1,\mathcal{S}_2),\alpha\in\boldsymbol{F},\boldsymbol{x}\in\mathcal{S}_1,$$

并定义 $\mathcal{L}(\mathcal{S}_1,\mathcal{S}_2)$ 中两个变换 T_1 与 T_2 相等,记为 $T_1=T_2$,假如 $T_1(\boldsymbol{x})=T_2(\boldsymbol{x})$, $\forall\,\boldsymbol{x}\in\mathcal{S}_1$,定义 $\mathcal{L}(\mathcal{S}_1,\mathcal{S}_2)$ 中零变换为 $O:O(\boldsymbol{x})=\boldsymbol{0}$, $\forall\,\boldsymbol{x}\in\mathcal{S}_1$,则在加法与数乘运算下, $\mathcal{L}(\mathcal{S}_1,\mathcal{S}_2)$ 组成域 $\boldsymbol{F}$ 上的线性空间,其维数等于 mn.

现考虑 $\boldsymbol{F}$ 上三个有限维线性空间 $\mathcal{S}_1,\mathcal{S}_2$ 与 $\mathcal{S}_3$. 定义 $T_2\in\mathcal{L}(\mathcal{S}_2,\mathcal{S}_3)$ 与 $T_1\in\mathcal{L}(\mathcal{S}_1,\mathcal{S}_2)$ 的**乘积或复合**(composition)为

$$(T_2T_1)(\boldsymbol{x})\equiv T_2(T_1(\boldsymbol{x})),\forall\,\boldsymbol{x}\in\mathcal{S}_1.$$

这时, $T=T_2T_1\in\mathcal{L}(\mathcal{S}_1,\mathcal{S}_3)$,并且,只要下列变换乘积有定义,则

$$T_3(T_2T_1)=(T_3T_2)T_1,$$

即变换乘积为可结合的. 容易看出,当 $\mathcal{S}_1=\mathcal{S}_2=\mathcal{S}$ 时,线性空间 $\mathcal{L}(\mathcal{S})$ 在上述复合运算下组成域 $\boldsymbol{F}$ 上的一个结合代数. 它以恒等变换 $I:I(\boldsymbol{x})\equiv\boldsymbol{x}$, $\forall\,\boldsymbol{x}\in\mathcal{S}$ 为单位元,即有 $IT=TI=T$, $\forall\,T\in\mathcal{L}(\mathcal{S})$.

$\mathcal{L}(\mathcal{S}_1,\mathcal{S}_2)$ 内的线性变换可由矩阵来表示. 设 $\mathscr{E}=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n\}$ 为 $\mathcal{S}_1$ 的基. 若 $\boldsymbol{x}\in\mathcal{S}_1$,则 $\boldsymbol{x}=\alpha_1\boldsymbol{x}_1+\alpha_2\boldsymbol{x}_2+\cdots+\alpha_n\boldsymbol{x}_n$,其中 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 为 $\boldsymbol{x}$ 关于基 $\mathscr{E}$ 的坐标. 记

$[\boldsymbol{x}]_{\mathscr{E}}=(\alpha_1,\alpha_2,\cdots,\alpha_n)^{\mathrm{T}}$. 显然，对应 $\boldsymbol{x}\to[\boldsymbol{x}]_{\mathscr{E}}$ 为（由 $\mathscr{S}_1$ 至 $\boldsymbol{F}^n$ 上的）同构. 通常$[\boldsymbol{x}]_{\mathscr{E}}\in\boldsymbol{F}^n$ 称为 $\boldsymbol{x}\in\mathscr{S}_1$ 关于基 $\mathscr{E}$ 的**表示**. 现设 $T\in\mathscr{L}(\mathscr{S}_1,\mathscr{S}_2)$，$\mathscr{S}_2$ 有基 $\mathscr{G}=\{\boldsymbol{y}_1,\boldsymbol{y}_2,\cdots,\boldsymbol{y}_m\}$. 则有 $T\boldsymbol{x}=T(\sum_{j=1}^{n}\alpha_j\boldsymbol{x}_j)=\sum_{j=1}^{n}\alpha_jT(\boldsymbol{x}_j)$，也就是说，一旦知道$[\boldsymbol{x}]_{\mathscr{E}}$，$T\boldsymbol{x}$ 便由 $T(\boldsymbol{x}_j)$ $(1\leqslant j\leqslant n)$完全确定. 假如$[T(\boldsymbol{x}_j)]_{\mathscr{G}}=(a_{1j},a_{2j},\cdots,a_{mj})^{\mathrm{T}}\in\boldsymbol{F}^m$，$j=1,2,\cdots,n$，那么，对任意 $\boldsymbol{x}\in\mathscr{S}_1$，

$$[T(\boldsymbol{x})]_{\mathscr{G}}=\Big[\sum_{j=1}^{n}\alpha_jT(\boldsymbol{x}_j)\Big]_{\mathscr{G}}=\sum_{j=1}^{n}\alpha_j[T(\boldsymbol{x}_j)]_{\mathscr{G}}$$

$$=\sum_{j=1}^{n}\alpha_j\begin{bmatrix}a_{1j}\\a_{2j}\\\vdots\\a_{mj}\end{bmatrix}=\begin{bmatrix}a_{11}&a_{12}&\cdots&a_{1n}\\a_{21}&a_{22}&\cdots&a_{2n}\\\vdots&\vdots&&\vdots\\a_{m1}&a_{m2}&\cdots&a_{mn}\end{bmatrix}\begin{bmatrix}\alpha_1\\\alpha_2\\\vdots\\\alpha_n\end{bmatrix}. \tag{1}$$

这里，$A\equiv(a_{ij})\in\boldsymbol{M}_{m,n}(\boldsymbol{F})$依赖于 T 与基对$(\mathscr{E},\mathscr{G})$的选取，称之为变换 T 关于基对$(\mathscr{E},\mathscr{G})$的**(矩阵)表示**，记为${}_{\mathscr{G}}[T]_{\mathscr{E}}$. 因此，(1)式可改写为

$$[T(\boldsymbol{x})]_{\mathscr{G}}={}_{\mathscr{G}}[T]_{\mathscr{E}}[\boldsymbol{x}]_{\mathscr{E}},\qquad\forall\,\boldsymbol{x}\in\mathscr{S}_1. \tag{2}$$

显然，同一变换在不同基对下一般有不同的(矩阵)表示. 另一方面，若 $\mathscr{S}_1,\mathscr{S}_2,\mathscr{E}$ 与 $\mathscr{G}$ 如前所述，给定 $A=(a_{ij})\in\boldsymbol{M}_{m,n}(\boldsymbol{F})$，则存在 $\mathscr{L}(\mathscr{S}_1,\mathscr{S}_2)$的唯一变换 T_A，它关于基对$(\mathscr{E},\mathscr{G})$的表示${}_{\mathscr{G}}[T_A]_{\mathscr{E}}$ 恰为 A. 事实上，对任意 $\boldsymbol{x}=\sum_{j=1}^{n}\alpha_j\boldsymbol{x}_j$，定义 $T_A(\boldsymbol{x})\equiv\sum_{j=1}^{n}\alpha_jT_A(\boldsymbol{x}_j)$，其中，$T_A(\boldsymbol{x}_j)=\sum_{i=1}^{m}a_{ij}\boldsymbol{y}_i$，$j=1,\cdots,n$. 容易验明，$T_A\in\mathscr{L}(\mathscr{S}_1,\mathscr{S}_2)$满足上述要求. 因此，当给定 $\mathscr{S}_1$ 与 $\mathscr{S}_2$ 的基对$(\mathscr{E},\mathscr{G})$时，线性空间 $\mathscr{L}(\mathscr{S}_1,\mathscr{S}_2)$与$\boldsymbol{M}_{m,n}(\boldsymbol{F})$ 之间存在一一对应：$T\leftrightarrow{}_{\mathscr{G}}[T]_{\mathscr{E}}$. 并且，若 $T_1\leftrightarrow{}_{\mathscr{G}}[T_1]_{\mathscr{E}}$ 与 $T_2\leftrightarrow{}_{\mathscr{G}}[T_2]_{\mathscr{E}}$，则有 $\alpha T_1+\beta T_2\leftrightarrow\alpha{}_{\mathscr{G}}[T_1]_{\mathscr{E}}+\beta{}_{\mathscr{G}}[T_2]_{\mathscr{E}}$，$\forall\,\alpha,\beta\in\boldsymbol{F}$，即对应 $T\leftrightarrow{}_{\mathscr{G}}[T]_{\mathscr{E}}$ 为两个线性空间 $\mathscr{L}(\mathscr{S}_1,\mathscr{S}_2)$与 $\boldsymbol{M}_{m,n}(\boldsymbol{F})$之间的同构.

现在考虑矩阵的乘法运算与线性变换复合运算之间的关系. 设 $T_1\in\mathscr{L}(\mathscr{S}_1,\mathscr{S}_2)$与 $T_2\in\mathscr{L}(\mathscr{S}_2,\mathscr{S}_3)$，$\mathscr{E}$ 与 $\mathscr{G}$ 含义如前，$\mathscr{R}=\{\boldsymbol{z}_1,\boldsymbol{z}_2,\cdots,\boldsymbol{z}_l\}$为 l 维线性空间 $\mathscr{S}_3$ 的基. 若 $A=(a_{ij})_{i,j=1}^{m;n}={}_{\mathscr{G}}[T_1]_{\mathscr{E}}$ 与 $B=(b_{ij})_{i,j=1}^{l;m}={}_{\mathscr{R}}[T_2]_{\mathscr{G}}$，则直接验证便知，${}_{\mathscr{R}}[T_2\,T_1]_{\mathscr{E}}=BA={}_{\mathscr{R}}[T_2]_{\mathscr{G}\mathscr{G}}[T_1]_{\mathscr{E}}$.

当 $\mathscr{S}_1=\mathscr{S}_2=\mathscr{S}_3=\mathscr{S}$时，自然地可以选取 $\mathscr{E}=\mathscr{G}=\mathscr{R}$，这时上述讨论仍然有效，不过，线性变换 $T\in\mathscr{L}(\mathscr{S})$的矩阵表示应为 $\boldsymbol{M}_n(\boldsymbol{F})$的矩阵$[T]_{\mathscr{E}}\equiv{}_{\mathscr{E}}[T]_{\mathscr{E}}$. 这时，$T\leftrightarrow[T]_{\mathscr{E}}$ 确定出代数 $\mathscr{L}(\mathscr{S})$与代数 $\boldsymbol{M}_n(\boldsymbol{F})$之间的同构.

同一变换 $T\in\mathscr{L}(\mathscr{S}_1,\mathscr{S}_2)$关于不同基对$(\mathscr{E},\mathscr{G})$与$(\mathscr{E}',\mathscr{G}')$的矩阵表示之间有着如下基本关系：

$${}_{\mathscr{G}'}[T]_{\mathscr{E}'}=P{}_{\mathscr{G}}[T]_{\mathscr{E}}Q^{-1}, \tag{3}$$

这里,$P\in \boldsymbol{M}_m(\boldsymbol{F})$与$Q\in \boldsymbol{M}_n(\boldsymbol{F})$分别表示由$\mathscr{G}$到$\mathscr{G}'$与由$\mathscr{E}$至$\mathscr{E}'$的转移矩阵,即$P$与$Q$满足

$$_{\mathscr{G}'}[T]_{\mathscr{E}} = P_{\mathscr{G}}[T]_{\mathscr{E}}, \qquad _{\mathscr{G}'}[T]_{\mathscr{E}} = {}_{\mathscr{G}'}[T]_{\mathscr{E}'}Q. \tag{4}$$

我们称$\boldsymbol{M}_{m,n}(\boldsymbol{F})$中两个矩阵$A$与$B$是**等价的**,记为$A\sim B$,假如有非奇异矩阵$E$与$F$使得$A=EBF$. 由于任一非奇异矩阵可分解为初等变换矩阵之积,故$A\sim B$等同于经过有限步关于行与列的矩阵初等变换,A可以变到B,反之,B可以变到A. 这样一来,(3)式表明同一变换$T\in\mathscr{L}(\mathscr{S}_1,\mathscr{S}_2)$关于不同基对的矩阵表示是等价的. 反之,若$A\in\boldsymbol{M}_{m,n}(\boldsymbol{F})$为$T\in\mathscr{L}(\mathscr{S}_1,\mathscr{S}_2)$的一个矩阵表示,则与$A$等价的矩阵$B\in\boldsymbol{M}_{m,n}(\boldsymbol{F})$也是$T$关于某基对的矩阵表示. 今后用$\mathscr{A}_T$表示$T\in\mathscr{L}(\mathscr{S}_1,\mathscr{S}_2)$的所有矩阵表示的集合. 显然,$\mathscr{A}_T\subset\boldsymbol{M}_{m,n}(\boldsymbol{F})$,$\mathscr{A}_T$内任意两个矩阵彼此等价,反之,$\boldsymbol{M}_{m,n}(\boldsymbol{F})$的任一给定的等价类对应着唯一的$T\in\mathscr{L}(\mathscr{S}_1,\mathscr{S}_2)$. 因此,对应$T\leftrightarrow\mathscr{A}_T$是一一的. 自然地,我们希望从$\mathscr{A}_T$中找出形式"最简单"的矩阵,作为$T$的标准表示. 由(3)式看出,当$T\neq O$时,$\mathscr{A}_T$中最简单矩阵可以取为下列形式之一:

$$\begin{bmatrix} I_r & O \\ O & O \end{bmatrix}, [I_m \quad O], \begin{bmatrix} I_n \\ O \end{bmatrix}, I_n, \tag{5}$$

它们分别对应于$\mathscr{A}_T$中矩阵的秩等于r、行满秩、列满秩与满秩($m=n$)的情形. (5)式中任一种矩阵可能作为**关于等价意义下T表示的标准型**.

对于$\mathscr{L}(\mathscr{S})$中的变换T,取$\mathscr{E}=\mathscr{G}$与$\mathscr{E}'=\mathscr{G}'$,在这个附加条件之下,一般不能得到关于等价意义下$T$的表示的诸如(5)式的标准型. 但按(3)与(4)两式,我们有$P=Q$,因此,当P为由$\mathscr{E}$到$\mathscr{E}'$的转移矩阵时,

$$[T]_{\mathscr{E}'} = P[T]_{\mathscr{E}}P^{-1}. \tag{6}$$

这个由(6)式确定的$\boldsymbol{M}_n(\boldsymbol{F})$中两个矩阵$[T]_{\mathscr{E}}$与$[T]_{\mathscr{E}'}$之间的关系便为矩阵论中最基本的关系之一:**相似**. 一般地说,两个同阶矩阵A与B称为**相似的**,记为$A\approx B$,假如对某非奇异矩阵P有

$$A = PBP^{-1}. \tag{7}$$

容易验证,相似为一种等价关系(即满足自反性、对称性与传递性的关系). 今后将同一变换$T\in\mathscr{L}(\mathscr{S})$关于$\mathscr{S}$内不同基的表示的集合记为$\mathscr{A}_T^0$. (6)式表明,$\mathscr{A}_T^0$是$\boldsymbol{M}_n(\boldsymbol{F})$中一个相似矩阵的等价类. 我们将在本小节后面讨论关于相似意义下$T\in\mathscr{L}(\mathscr{S})$的表示的标准型,即Jordan正规形式问题.

按1.1.1小节的结果,若$A\approx B$,则有$\operatorname{rank}A=\operatorname{rank}B$, $\det A=\det B$, $A-\lambda I\approx B-\lambda I$, $\forall\lambda\in\boldsymbol{F}$, $A^k\approx B^k(k=1,2,\cdots)$, $\operatorname{tr}A=\operatorname{tr}B$, $A^{\mathrm{T}}\approx B^{\mathrm{T}}$,此外,$A$与$B$有相同的特征值(计入重数).

类似于矩阵的情形,可以考虑变换$T\in\mathscr{L}(\mathscr{S}_1,\mathscr{S}_2)$的**值域**$\operatorname{Im}T$与**核**$\operatorname{Ker}T$:

$$\operatorname{Im}T \equiv \{T(\boldsymbol{x}):\boldsymbol{x}\in\mathscr{S}_1\} = T(\mathscr{S}_1)\subset\mathscr{S}_2,$$
$$\operatorname{Ker}T \equiv \{\boldsymbol{x}\in\mathscr{S}_1:T(\boldsymbol{x})=\boldsymbol{0}\}\subset\mathscr{S}_1.$$

它们分别为 $\mathscr{S}_2$ 与 $\mathscr{S}_1$ 的线性子空间. $T\in\mathscr{L}(\mathscr{S}_1,\mathscr{S}_2)$的**秩**由

$$\text{rank}T\equiv\dim(\text{Im}T) \tag{8}$$

来定义;它的**零度**由

$$\text{nul}T\equiv\dim(\text{Ker}T) \tag{9}$$

来定义. 可以证明(见习题 1):

$$\text{rank}T=\text{rank}A,\qquad \text{nul}T=\text{nul}A, \forall A\in\mathscr{A}_T. \tag{10}$$

因此,按 1.1.1 小节定理 6,我们有

$$\text{rank}T+\text{nul}T=\dim\mathscr{S}_1,\qquad \forall T\in\mathscr{L}(\mathscr{S}_1,\mathscr{S}_2). \tag{11}$$

有时将 $\dim\mathscr{S}_2$ 与 $\text{rank}T$ 之差叫作 $T\in\mathscr{L}(\mathscr{S}_1,\mathscr{S}_2)$ 的**亏量**(deficiency),记为 $\text{def}T$. 因此,

$$\text{rank}T+\text{def}T=\dim\mathscr{S}_2,\qquad \forall T\in\mathscr{L}(\mathscr{S}_1,\mathscr{S}_2). \tag{12}$$

特殊地,对于变换 $T\in\mathscr{L}(\mathscr{S})(\dim\mathscr{S}=n)$,按定义有

$$\begin{gathered}\text{Im}T^0\supset\text{Im}T\supset\text{Im}T^2\supset\text{Im}T^3\supset\cdots,\\ \text{Ker}T^0\subset\text{Ker}T\subset\text{Ker}T^2\subset\text{Ker}T^3\subset\cdots,\end{gathered} \tag{13}$$

并存在最小非负整数 k 使得 $\text{Im}T^l=\text{Im}T^k$ 与 $\text{Ker}T^l=\text{Ker}T^k$, $\forall\, l\geqslant k$,因为(13)式中所有集合均包含在 $\mathscr{S}$内.

现在考虑线性变换的逆变换. 设 $T\in\mathscr{L}(\mathscr{S}_1,\mathscr{S}_2)$, $\dim\mathscr{S}_1=n$ 与 $\dim\mathscr{S}_2=m$. T 称为**左可逆的**,假如有 $T_1\in\mathscr{L}(\mathscr{S}_2,\mathscr{S}_1)$使得 $T_1T=I_1$,这里 I_1 为 $\mathscr{S}_1$ 上恒等变换,这时 T_1 叫作 T 的一个**左逆**. 类似地,T 称为**右可逆的**,假如有 $T_2\in\mathscr{L}(\mathscr{S}_2,\mathscr{S}_1)$使得 $TT_2=I_2$,这里 I_2 为 $\mathscr{S}_2$ 上恒等变换,这时 T_2 叫作 T 的一个**右逆**. 左可逆或右可逆线性变换有如下基本特征,其证明留给读者(见习题 2).

定理 1　设 $T\in\mathscr{L}(\mathscr{S}_1,\mathscr{S}_2)$, $\dim\mathscr{S}_1=n$, $\dim\mathscr{S}_2=m$. 则下列诸命题等价:

(1) T 为左(右)可逆的.

(2) $\text{Ker}T=\{\mathbf{0}\}$($\text{rank}T=m$).

(3) T 确定 $\mathscr{S}_1$ 与 $\text{Im}T$($\text{Ker}T$ 的任意直补与 $\mathscr{S}_2$)之间的一个同构.

(4) $n\leqslant m$ 与 $\text{rank}T=n$($n\geqslant m$ 与 $\text{nul}T=n-m$).

(5) 任意 $A\in\mathscr{A}_T$ 为 $\boldsymbol{M}_{m,n}(\boldsymbol{F})$中列(行)满秩矩阵.

(6) 任意 $A\in\mathscr{A}_T$ 为 $\boldsymbol{M}_{m,n}(\boldsymbol{F})$中左(右)可逆矩阵,即存在 $B\in\boldsymbol{M}_{n,m}(\boldsymbol{F})$使得 $BA=I_n$($AB=I_m$).

变换 $T\in\mathscr{L}(\mathscr{S}_1,\mathscr{S}_2)$称为**可逆的**,假如 T 同时是左与右可逆的. 因此,由定理 1 直接推出可逆变换的如下特征.

定理 2　设 $T\in\mathscr{L}(\mathscr{S}_1,\mathscr{S}_2)$, $\dim\mathscr{S}_1=n$, $\dim\mathscr{S}_2=m$. 则下列诸命题等价:

(1) T 为可逆的.

(2) $\text{Ker}T=\{\mathbf{0}\}$与 $\text{rank}T=m=n$.

(3) T 确定 $\mathscr{S}_1$ 与 $\mathscr{S}_2$ 之间的一个同构.

(4) 任意 $A\in\mathscr{A}_T$ 为 $\boldsymbol{M}_n(\boldsymbol{F})$ 中非奇异矩阵.

可逆变换 $T\in\mathscr{L}(\mathscr{S}_1,\mathscr{S}_2)$ 存在唯一的**逆变换**,记为 T^{-1},满足 $T^{-1}T=I_1$ 与 $TT^{-1}=I_2$. 若 T 关于基对$(\mathscr{E},\mathscr{G})$的表示为 $A\in\boldsymbol{M}_n(\boldsymbol{F})$,则 $T^{-1}\in\mathscr{L}(\mathscr{S}_2,\mathscr{S}_1)$关于基对$(\mathscr{G},\mathscr{E})$的表示恰为 A^{-1}. 特殊地,若 $T,T^{-1}\in\mathscr{L}(\mathscr{S})$, $\dim\mathscr{S}=n$,则有 $T^{-1}T=TT^{-1}=I$($\mathscr{S}$上恒等变换),且若 T 关于基 $\mathscr{E}$ 的表示为 A,则 T^{-1}关于基 $\mathscr{E}$ 的表示恰为 A^{-1}.

对于 $T\in\mathscr{L}(\mathscr{S}_1,\mathscr{S}_2)$与 $\mathscr{S}_1$ 的线性子空间 $\mathscr{S}_0$,我们可考虑变换 T 对于 $\mathscr{S}_0$ 的**限制**(restriction),记为 $T|_{\mathscr{S}_0}$,它由下式定义:

$$T|_{\mathscr{S}_0}(\boldsymbol{x})=T(\boldsymbol{x}),\ \forall\,\boldsymbol{x}\in\mathscr{S}_0. \tag{14}$$

显然,$\mathrm{Im}(T|_{\mathscr{S}_0})=T(\mathscr{S}_0)$与 $\mathrm{Ker}(T|_{\mathscr{S}_0})=\mathscr{S}_0\cap\mathrm{Ker}T$. 现假定 $\mathscr{S}_1$ 可分解为两个子空间 $\mathscr{S}_0$ 与 $\widetilde{\mathscr{S}}$的直和:$\mathscr{S}_1=\mathscr{S}_0\dotplus\widetilde{\mathscr{S}}$(即有 $\mathscr{S}_1=\widetilde{\mathscr{S}}_0+\widetilde{\mathscr{S}},\mathscr{S}_0\cap\widetilde{\mathscr{S}}=\{\boldsymbol{0}\}$). 对 $\boldsymbol{x}\in\mathscr{S}_1$,记 $\boldsymbol{x}=\boldsymbol{x}_1+\boldsymbol{x}_2$,其中 $\boldsymbol{x}_1\in\mathscr{S}_0,\boldsymbol{x}_2\in\widetilde{\mathscr{S}}$. 当 $T_1\in\mathscr{L}(\mathscr{S}_0,\mathscr{S}_2)$与 $T_2\in\mathscr{L}(\widetilde{\mathscr{S}},\mathscr{S}_2)$时,下式

$$T(\boldsymbol{x})=T_1(\boldsymbol{x}_1)+T_2(\boldsymbol{x}_2),\qquad\forall\,\boldsymbol{x}\in\mathscr{S}_1 \tag{15}$$

成立当且仅当 $T_1=T|_{\mathscr{S}_0}$与 $T_2=T|_{\widetilde{\mathscr{S}}}$. 类似的结论对 $\mathscr{S}_1$ 分解为多个子空间直和的情形也成立.

当限定 $\mathscr{S}_1=\mathscr{S}_2=\mathscr{S}$时,$\mathscr{S}$的子空间 $\mathscr{S}_0$ 称为在变换 $T\in\mathscr{L}(\mathscr{S})$下**不变的**(或 **$T$-不变的**),假如 $T(\mathscr{S}_0)\subset\mathscr{S}_0$. 此时,$T|_{\mathscr{S}_0}$便是 $\mathscr{S}_0$ 上线性变换. 因为 $\mathrm{Im}(T|_{\mathscr{S}_0})=T(\mathscr{S}_0)\subset\mathscr{S}_0$,所以它可视为 T 在 $\mathscr{S}_0$ 内的“部分变换”.

对 $T\in\mathscr{L}(\mathscr{S})$来说,$\{\boldsymbol{0}\}$与 $\mathscr{S}$必定为 T-不变子空间,它们为平凡情形. 对于不可逆非零变换 $T\in\mathscr{L}(\mathscr{S})$按定理 2 总有非平凡的 T-不变子空间:$\mathrm{Im}T$ 与 $\mathrm{Ker}T$.

对于任意矩阵 $A=(a_{ij})\in\boldsymbol{M}_n(\boldsymbol{F})$,它的不变子空间可定义为与 A 对应的线性变换 $T_A\in\mathscr{L}(\boldsymbol{F}^n)$的不变子空间,这里,

$$T_A(\boldsymbol{x})=A\boldsymbol{x},\quad\forall\,\boldsymbol{x}\in\boldsymbol{F}^n. \tag{16}$$

假如 $\mathscr{S}_0$ 为 k 维 T-不变子空间,则存在 $A\in\mathscr{A}_T^0$,它也有一个 k 维不变子空间. 事实上,设 $\mathscr{E}_0=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_k\}$为 $\mathscr{S}_0$ 的基,$\mathscr{E}=\{\boldsymbol{x}_1,\cdots,\boldsymbol{x}_k,\boldsymbol{x}_{k+1},\cdots,\boldsymbol{x}_n\}$为 $\mathscr{S}$的基. 若 $\boldsymbol{y}=T(\boldsymbol{x}),\boldsymbol{x}=\sum\limits_{i=1}^{k}\alpha_i\boldsymbol{x}_i\in\mathscr{S}_0$,则由于 $\mathscr{S}_0$ 为 T-不变的,$[\boldsymbol{y}]_{\mathscr{E}}=(\beta_1,\cdots,\beta_k,0,\cdots,0)^{\mathrm{T}}\in\boldsymbol{F}^n$. 另一方面,我们有$[\boldsymbol{x}]_{\mathscr{E}}=(\alpha_1,\cdots,\alpha_k,0,\cdots,0)^{\mathrm{T}}\in\boldsymbol{F}^n$. 按前面讨论,$\boldsymbol{y}=T(\boldsymbol{x})$蕴涵$[\boldsymbol{y}]_{\mathscr{E}}=[T]_{\mathscr{E}}[\boldsymbol{x}]_{\mathscr{E}}$,这表示维数为 k 的 $\boldsymbol{F}^n$ 子空间$\{[\boldsymbol{x}]_{\mathscr{E}}:\boldsymbol{x}\in\mathscr{S}_0\}$是$[T]_{\mathscr{E}}\in\mathscr{A}_T^0$的不变子空间.

对于 $T\in\mathscr{L}(\mathscr{S})$,为寻求相似矩阵等价类 $\mathscr{A}_T^0$的标准型,首先要讨论线性变换与矩阵的可约性与完全可约性问题. 假如 $\mathscr{S}=\mathscr{S}_1\dotplus\mathscr{S}_2$,其中,$\mathscr{S}_1$ 与 $\mathscr{S}_2$ 至少有一个为 T-不变的,不妨设 $T\mathscr{S}_1\subset\mathscr{S}_1$, $\dim\mathscr{S}_1=k$. 我们选取 $\mathscr{S}$的基 $\mathscr{E}=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n\}$使得前面 k 个元素组成 $\mathscr{S}_1$ 的基 $\mathscr{E}_1$,余下 $n-k$ 个元素组成 $\mathscr{S}_2$ 的基 $\mathscr{E}_2$,则因 $\mathscr{S}_1$ 为 T-不变的,$T(\boldsymbol{x}_j)\subset\mathscr{S}_1(1\leqslant j\leqslant k)$,这时,我们有

$$T(\boldsymbol{x}_j)=\begin{cases}\sum_{i=1}^{k}a_{ij}\boldsymbol{x}_i, & j=1,\cdots,k,\\ \sum_{i=1}^{n}a_{ij}\boldsymbol{x}_i, & j=k+1,\cdots,n.\end{cases}\tag{17}$$

因此，$[T]_{\mathscr{E}}=(a_{ij})\in\boldsymbol{M}_n(\boldsymbol{F})$具有形式

$$[T]_{\mathscr{E}}=\begin{bmatrix}A_{11} & A_{12}\\ O & A_{22}\end{bmatrix},\qquad A_{11}\in\boldsymbol{M}_k(\boldsymbol{F}),\tag{18}$$

其中，A_{11}为$T_1\equiv T|_{\mathscr{S}_1}$（关于基$\mathscr{E}_1$）的矩阵表示. 这个事实表明，当$\mathscr{S}$满足前述特殊分解条件时，只要取$\mathscr{S}$的合适基$\mathscr{E}$，$T$就有分块三角矩阵表示(18). 假如$\mathscr{S}_2$也是$T$-不变的，按前面推导，$[T]_{\mathscr{E}}=(a_{ij})\in\boldsymbol{M}_n(\boldsymbol{F})$具有分块对角矩阵形式

$$[T]_{\mathscr{E}}=\begin{bmatrix}A_{11} & O\\ O & A_{22}\end{bmatrix},\quad A_{11}\in\boldsymbol{M}_k(\boldsymbol{F}).\tag{19}$$

这时，对每一个$\boldsymbol{x}=\boldsymbol{x}^1+\boldsymbol{x}^2$，$\boldsymbol{x}^1\in\mathscr{S}_1$与$\boldsymbol{x}^2\in\mathscr{S}_2$，

$$T(\boldsymbol{x})=T_1(\boldsymbol{x}^1)+T_2(\boldsymbol{x}^2),\tag{20}$$

式中，$T_i=T|_{\mathscr{S}_i}$，$i=1,2$，自然地，我们将$T\in\mathscr{L}(\mathscr{S})$叫作**线性变换$T_1$与$T_2$的直和**，记为$T=T_1\dotplus T_2$. 容易看到(19)式中$A_{11}$与$A_{22}$分别为$T_1\in\mathscr{L}(\mathscr{S}_1)$与$T_2\in\mathscr{L}(\mathscr{S}_2)$的矩阵表示. 反之，假如$\mathscr{A}_T^0$中有形如(19)式的矩阵，则$\mathscr{S}$显然有基$\mathscr{E}=\{\boldsymbol{x}_1,\cdots,\boldsymbol{x}_k,\boldsymbol{x}_{k+1},\cdots,\boldsymbol{x}_n\}$使得下式成立：

$$T(\boldsymbol{x}_j)=\begin{cases}\sum_{i=1}^{k}a_{ij}\boldsymbol{x}_i, & j=1,\cdots,k,\\ \sum_{i=k+1}^{n}a_{ij}\boldsymbol{x}_i, & j=k+1,\cdots,n.\end{cases}\tag{21}$$

于是，由$\boldsymbol{x}_1,\cdots,\boldsymbol{x}_k$生成的$\mathscr{S}$的子空间$\mathscr{S}_1$与由$\boldsymbol{x}_{k+1},\cdots,\boldsymbol{x}_n$生成的$\mathscr{S}$的子空间$\mathscr{S}_2$均为$T$-不变的，且$\mathscr{S}=\mathscr{S}_1\dotplus\mathscr{S}_2$. 与变换直和概念相对应，通常将形如(19)式的矩阵称为A_{11}与A_{22}的**直和**，记为$A_{11}\dotplus A_{22}$.

容易将上述结果推广到多个线性变换或矩阵直和的情形，即$T=T_1\dotplus\cdots\dotplus T_p\equiv\dot{\sum\limits_{i=1}^{p}}T_i$或$A\equiv\dot{\sum\limits_{i=1}^{p}}A_{ii}$的情形.

类似于方阵情形，我们可以考虑变换$T\in\mathscr{L}(\mathscr{S})$的特征值与特征向量. 若有$\boldsymbol{F}=\mathbb{C}$且

$$T\boldsymbol{x}=\lambda\boldsymbol{x},\quad \boldsymbol{x}\neq\boldsymbol{0},\tag{22}$$

则称λ为T的**特征值**，$\boldsymbol{x}\in\mathscr{S}$为对应于$\lambda$的$T$的**特征向量**. 取定$\mathscr{S}$的基$\mathscr{E}$，按1.1.1小节讨论我们有

$$[T\boldsymbol{x}]_{\mathscr{E}}=[T]_{\mathscr{E}}[\boldsymbol{x}]_{\mathscr{E}}=\lambda[\boldsymbol{x}]_{\mathscr{E}},\tag{23}$$

因而λ为T关于基$\mathscr{E}$矩阵表示$[T]_{\mathscr{E}}$的特征值，$\boldsymbol{x}$关于基的表示$[\boldsymbol{x}]_{\mathscr{E}}$为对应于$\lambda$的

$[T]_{\mathscr{E}}$ 的特征向量. 因此,T 与$[T]_{\mathscr{E}}$ 有相同的特征值. 若 λ 为 $T\in\mathscr{L}(\mathscr{S})$的特征值,则 $\mathrm{Ker}(T-\lambda I)$叫作对应于 λ 的 **T 的特征空间**. 显然,T 的特征空间内的任意非零向量为 T 的特征向量.

为了寻求 $\mathscr{A}_T^0$中标准型,我们还需要考虑 $T\in\mathscr{L}(\mathscr{S})$的**零化多项式**与**最小多项式**的概念. 注意,在本小节余下部分总假定 $\boldsymbol{F}=\mathbb{C}$,它保证 $\boldsymbol{F}$ 为代数闭域.

一个非零数值多项式 $p(\lambda)=\sum\limits_{i=0}^{l}p_i\lambda^i(p_l\neq 0)$叫作变换 $T\in\mathscr{L}(\mathscr{S})(\dim\mathscr{S}=n)$ 的**零化多项式**,假如 $p(T)=\sum\limits_{i=0}^{l}p_iT^i=O$. 由于 $\mathscr{L}(\mathscr{S})$为有限维的$(\dim\mathscr{L}(\mathscr{S})=n^2)$,故只要取 l 充分大,就有 $I,T,\cdots,T^l$ 线性相关,于是有 $\sum\limits_{i=0}^{l}p_iT^i=O$,其中$p_0,\cdots,p_l$ 不全为零,这证明 T 的零化多项式的存在性. 显然,T 的零化多项式不是唯一的. 若 $T\in\mathscr{L}(\mathscr{S})$的某个零化多项式为 $p(\lambda)=\prod\limits_{i=1}^{k}p_i(\lambda)$,其中,$p_1(\lambda),\cdots,p_k(\lambda)$为两两互素多项式,则有

$$\mathscr{S}=\mathrm{Ker}(p(T))=\dot{\sum_{i=1}^{k}}\mathrm{Ker}(p_i(T)). \tag{24}$$

例如,对于下列矩阵(视为$\mathbb{C}^4$ 上的线性变换)

$$A=\begin{bmatrix}\lambda_1 & 1\\ 0 & \lambda_1\end{bmatrix}\dot{+}\begin{bmatrix}\lambda_2 & 1\\ 0 & \lambda_2\end{bmatrix},\qquad \lambda_1\neq\lambda_2, \tag{25}$$

$p(\lambda)=(\lambda-\lambda_1)^2(\lambda-\lambda_2)^2$ 为 A 的零化多项式. 令 $p_1(\lambda)=(\lambda-\lambda_1)^2,p_2(\lambda)=(\lambda-\lambda_2)^2$. 则 $\mathrm{Ker}(A-\lambda_1I_4)^2=\{(\alpha,\beta,0,0)^{\mathrm{T}}:\alpha,\beta\in\mathbb{C}\}$,$\mathrm{Ker}(A-\lambda_2I_4)^2=\{(0,0,\nu,\delta)^{\mathrm{T}}:\nu,\delta\in\mathbb{C}\}$,因而$\mathbb{C}^4=\mathrm{Ker}(A-\lambda_1I_4)^2\dot{+}\mathrm{Ker}(A-\lambda_2I_4)^2$,$\forall\lambda_1\neq\lambda_2$.

$T\in\mathscr{L}(\mathscr{S})$的零化多项式中次数最小的唯一的首一多项式称为 T 的**最小多项式**,记为 $m_T(\lambda)$或 $m(\lambda)$. T 的最小多项式 $m(\lambda)$整除 T 的任意零化多项式. T 的最小多项式 $m(\lambda)$的不同零点组成 T 的**谱**(即 T 的不同特征值集合)$\sigma(T)$(见习题3). 因此,若 $T\in\mathscr{L}(\mathscr{S})$的谱 $\sigma(T)=\{\lambda_1,\cdots,\lambda_s\}$,则 T 的最小多项式 $m(\lambda)$可以表示为

$$m(\lambda)=\prod_{i=1}^{s}(\lambda-\lambda_i)^{m_i}, \tag{26}$$

这里,诸 m_i 是与 $\lambda_i\in\sigma(T)$有关的唯一正整数,称为 λ_i 的**指标**(index),即 $m_i=\mathrm{index}\lambda_i$. 例如,(25)式中矩阵 A(视为$\mathbb{C}^4$ 上的线性变换)有 $m_A(\lambda)=p(\lambda)=(\lambda-\lambda_1)^2\cdot(\lambda-\lambda_2)^2$,$\lambda_1\neq\lambda_2$,因而 $\lambda_1,\lambda_2\in\sigma(A)$的指标均为 2. 根据(24)式,由(26)式推出

$$\mathscr{S}=\dot{\sum_{i=1}^{s}}\mathscr{S}_i, \tag{27}$$

这里,$\mathscr{S}_i=\mathrm{Ker}(T-\lambda_iI)^{m_i}$,$i=1,\cdots,s$. 由于 T 与 $T-\lambda_iI$ 可换,故各个 $\mathscr{S}_i$ 为 T-不变

子空间. 令 $T_i=T|_{\mathscr{S}_i},i=1,\cdots,s$,则$(\lambda-\lambda_i)^{m_i}$恰为 T_i 的最小多项式. 于是从$(T_i-\lambda_iI)^{m_i}=O,(T_i-\lambda_iI)^{m_i-1}\neq O$,以及下列严格包含关系

$$\mathrm{Ker}(T_i-\lambda_iI)\subset\mathrm{Ker}(T_i-\lambda_iI)^2\subset\cdots\subset\mathrm{Ker}(T_i-\lambda_iI)^{m_i} \tag{28}$$

可推出,$\dim\mathscr{S}_i=\dim(\mathrm{Ker}(T_i-\lambda_iI)^{m_i})\geqslant m_i,i=1,\cdots,s$. 按前面分析,这时有 T 的唯一分解(不计和项的次序)

$$T=\sum_{i=1}^{s}{}^{\cdot}\,T_i, \tag{29}$$

因而寻求 $T\in\mathscr{L}(\mathscr{S})$的标准表示可归结为寻求诸 T_i 的标准表示问题. 选取 $\mathscr{S}$的合适的基 $\mathscr{E}$,按前面分析我们有

$$[T]_{\mathscr{E}}=A=\mathrm{diag}[A_{11},A_{22},\cdots,A_{ss}], \tag{30}$$

这里,$\sigma(A_{ii})=\{\lambda_i\},i=1,\cdots,s$,且每个 A_{ii} 的阶数 q_i 由 T 唯一确定,$q_1+\cdots+q_s=n(=\dim\mathscr{S})$. 若对 $T\in\mathscr{L}(\mathscr{S})$,定义 $\det T=\det A,\ \forall A\in\mathscr{A}_T^0$,则由此可推出矩阵论中基本定理——**Cayley- Hamilton 定理**.

定理 3　$T\in\mathscr{L}(\mathscr{S})$的特征多项式 $c(\lambda)=\det(\lambda I-T)$为 T 的零化多项式,即有 $c(T)=O$.

证明　由于 $\det T=\det A,\ \forall A\in\mathscr{A}_T^0$,我们有 $c(\lambda)=\det(\lambda I_n-A),\ \forall A\in\mathscr{A}_T^0$. 按(30)式,$c(\lambda)=\prod\limits_{i=1}^{s}\det(\lambda I-A_{ii})$. 由于 $A_{ii}\in\boldsymbol{M}_{q_i}(\mathbb{C})$与 $\sigma(A_{ii})=\{\lambda_i\}$,故 $\det(\lambda I-A_{ii})=(\lambda-\lambda_i)^{q_i},i=1,\cdots,s$,且有 $q_i\geqslant m_i,i=1,\cdots,s$. 因此,$c(\lambda)$能被 T 的最小多项式 $m(\lambda)=\prod\limits_{i=1}^{s}(\lambda-\lambda_i)^{m_i}$整除,即有多项式 $d(\lambda)$使得

$$c(\lambda)=d(\lambda)m(\lambda),$$

由此推得 $c(T)=d(T)m(T)=O$. □

通常,把 $\mathscr{S}$的子空间 $\mathrm{Ker}(T-\lambda I)^r(1\leqslant r\leqslant\mathrm{index}\lambda,\lambda\in\sigma(T))$叫作对应于特征值 λ 的 T 的 **r 级广义特征空间**;把满足下列条件的非零元素 $\boldsymbol{x}\in\mathscr{S}$:

$$\boldsymbol{x}\in\mathrm{Ker}(T-\lambda I)^r\text{ 与 }\boldsymbol{x}\notin\mathrm{Ker}(T-\lambda I)^{r-1}, \tag{31}$$

叫作对应于特征值 λ 的 T 的 **r 级广义特征向量**或 **r 级主向量**. 特别地,对应于 λ 的 T 的特征空间与特征向量即为 T 的一级广义特征空间与一级广义特征向量. 按这些术语,(27)式中 $\mathscr{S}$可视为 s 个 T 的广义特征空间 $\mathscr{S}_i$ 的直和.

假如 $\boldsymbol{x}_r$ 为对应于$\lambda_i\in\sigma(T)$的 $T\in\mathscr{L}(\mathscr{S})$的 r 级广义特征向量,$r\geqslant2$,那么按定义,$\boldsymbol{x}_{r-1}\equiv(T-\lambda_iI)(\boldsymbol{x}_r)$便是 T 的 $r-1$ 级广义特征向量,因而

$$T(\boldsymbol{x}_r)=\lambda_i\boldsymbol{x}_r+\boldsymbol{x}_{r-1}. \tag{32}$$

反之,若 $\boldsymbol{x}_{r-1}=(T-\lambda_iI)(\boldsymbol{x}_r)$为对应于 λ_i 的 T 的 $r-1$ 级广义特征向量,则 $r\geqslant2$ 且 $\boldsymbol{x}_r$ 显然是对应于 λ_i 的 T 的 r 级广义特征向量. 按此步骤继续下去,可得 T 的广义特征向量列 $\boldsymbol{x}_r,\boldsymbol{x}_{r-1},\cdots,\boldsymbol{x}_2,\boldsymbol{x}_1$,使得

$$\begin{aligned} T(\boldsymbol{x}_1) &= \lambda_i \boldsymbol{x}_1, \\ T(\boldsymbol{x}_2) &= \lambda_i \boldsymbol{x}_2 + \boldsymbol{x}_1, \\ &\vdots \\ T(\boldsymbol{x}_r) &= \lambda_i \boldsymbol{x}_r + \boldsymbol{x}_{r-1}. \end{aligned} \tag{33}$$

上述序列 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_r$ 称为对应于 $\lambda_i \in \sigma(T)$ 的长度为 r 的 **Jordan 链**. 任意 Jordan 链中元素线性无关（见习题 4），于是，Jordan 链 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_r$ 生成的子空间 $\mathrm{span}\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_r\}$ 有维数 r，由(33)式看出，它还是 T-不变的，我们称之为 $T \in \mathscr{L}(\mathscr{S})$ 的 **Jordan 子空间**. 它仅含 T 的一个线性无关的特征向量. 因为 T 有 s 个不同特征值 $\lambda_1, \cdots, \lambda_s$，我们可以得到若干个 T 的 Jordan 子空间. 如果 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_r$ 为对应于 λ_i 的长度为 r 的 Jordan 链，则由(33)式看出，

$$\begin{aligned} \boldsymbol{x}_r &= (T - \lambda_i I)^0 (\boldsymbol{x}_r), \\ \boldsymbol{x}_{r-1} &= (T - \lambda_i I)(\boldsymbol{x}_r), \\ &\vdots \\ \boldsymbol{x}_2 &= (T - \lambda_i I)^{r-2} (\boldsymbol{x}_r), \\ \boldsymbol{x}_1 &= (T - \lambda_i I)^{r-1} (\boldsymbol{x}_r). \end{aligned} \tag{34}$$

应用二项式定理，由上式看出，$\mathrm{span}\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_r\} = \mathrm{span}\{\boldsymbol{x}_r, T(\boldsymbol{x}_r), \cdots, T^{r-1}(\boldsymbol{x}_r)\}$，也就是说，$T$ 的 Jordan 子空间 $\mathrm{span}\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_r\}$ 内有一元素 $\boldsymbol{x}_r$，使得 $\{\boldsymbol{x}_r, T(\boldsymbol{x}_r), \cdots, T^{r-1}(\boldsymbol{x}_r)\}$ 组成该子空间的基. 这时称这样的子空间为 $\mathscr{S}$ 的关于 T 的**循环子空间**. 此外，从(34)式知道，若 $\boldsymbol{x}_r$ 为对应于 λ_i 的 T 的 r 级广义特征向量，那么 $(T - \lambda_i I)^j \boldsymbol{x}_r$ 便为对应于 λ_i 的 T 的 $r-j$ 级广义特征向量 $(0 \leqslant j \leqslant r-1)$.

对于(27)式中 $\mathscr{S}_i = \mathrm{Ker}(T - \lambda_i I)^{m_i}$，$m_i = \mathrm{index}\lambda_i$，若不计诸和项前后次序，$\mathscr{S}_i$ 有唯一分解（见习题 5）：

$$\mathscr{S}_i = \dot{\sum_{j=1}^{t_{m_i}}} \mathscr{T}_{m_i}^{(j)} \dot{+} \dot{\sum_{j=1}^{t_{m_i-1}}} \mathscr{T}_{m_i-1}^{(j)} \dot{+} \cdots \dot{+} \dot{\sum_{j=1}^{t_1}} \mathscr{T}_1^{(j)}, \tag{35}$$

式中，$\mathscr{T}_k^{(j)}$ $(1 \leqslant k \leqslant m_i, 1 \leqslant j \leqslant t_k)$ 为维数等于 k 的对应于 $\lambda_i \in \sigma(T)$ 的关于 T 的循环 Jordan 子空间，其中某些 t_k 可能为零，这时相应的和项应理解为平凡子空间 $\{\boldsymbol{0}\}$. (35)式表明，在 $\mathscr{S}_i$ 中存在由若干个 Jordan 链拼接成的基，而每一个这样的 Jordan 链又是某循环 Jordan 子空间 $\mathscr{T}_k^{(j)}$ 的基. 通常，将这样给出的 $\mathscr{S}_i = \mathrm{Ker}(T - \lambda_i I)^{m_i}$ 的基称为此广义特征空间的 **Jordan 基**.

由于每一个 T 的 Jordan 子空间是 T-不变的，从(27)与(35)两式得出

$$T = \dot{\sum_{i=1}^{s}} T \mid \mathscr{S}_i = \dot{\sum_{i=1}^{s}} \dot{\sum_{i=1}^{t^{(i)}}} T_i^{(j)} \tag{36}$$

式中，$t^{(i)} = \dim(\mathrm{Ker}(T - \lambda_i I))$，而变换 $T_i^{(j)}$ $(1 \leqslant i \leqslant s, 1 \leqslant j \leqslant t^{(i)})$ 是变换 $T \in \mathscr{L}(\mathscr{S})$ 在对应于 λ_i 的 T 的某个循环 Jordan 子空间上的限制. 分解式(36)称为变换 $T \in \mathscr{L}(\mathscr{S})$ 的 **Jordan 分解**.

有了前面的讨论,我们现在可考虑 $T\in\mathscr{L}(\mathscr{S})$ 的矩阵表示的标准型问题. 由(36)式知道,实际上只要考虑变换 $T_i^{(j)}$ 的矩阵表示的标准型即可.

假定 $T_i^{(j)}$ 是 T 在由(34)式中 Jordan 链生成的 T 的循环 Jordan 子空间上的限制,其中,$\lambda_i\in\sigma(T)$,$\boldsymbol{x}_r$ 为对应于 λ_i 的 T 的 r 级广义特征向量,并且 $T_i^{(j)}$ 关于(34)式中基的矩阵表示为 $J_i^{(r)}=(\alpha_{pq})\in\boldsymbol{M}_r(\mathbb{C})$,即有 $T(\boldsymbol{x}_q)=\sum_{p=1}^{r}\alpha_{pq}\boldsymbol{x}_p(1\leqslant q\leqslant r)$. 比较(33)式,我们有

$$J_i^{(r)}=\begin{bmatrix}\lambda_i & 1 & 0 & \cdots & 0\\ 0 & \lambda_i & 1 & \cdots & 0\\ & & & & \vdots\\ \vdots & \vdots & \ddots & \ddots & 0\\ & & & & 1\\ 0 & 0 & \cdots & 0 & \lambda_i\end{bmatrix}\in\boldsymbol{M}_r(\mathbb{C}).\tag{37}$$

形如(37)式的矩阵 $J_i^{(r)}$ 称为对应于 $\lambda_i\in\sigma(T)$ 的 **r 阶 Jordan 块**,而 $(\lambda-\lambda_i)^r$ 叫作 $\lambda I_r-J_i^{(r)}$ 的**初等因子**,$i=1,\cdots,s;r=1,\cdots,m_i$. 对固定的 i,所有 $T_i^{(j)}$ 中 r 阶Jordan 块的个数记为 $t_r^{(i)}$,则有 $\sum_{r=1}^{m_i}t_r^{(i)}=t^{(i)}$. 在相似变换下,每个 Jordan 块均不能分解为更小的块.

所有 $\mathscr{S}_i(1\leqslant i\leqslant s)$ 的 Jordan 基拼接起来组成整个空间 $\mathscr{S}$ 的基,称之为 $\mathscr{S}$ 的 **T-Jordan 基**. 应用(35)式与(36)式以及(37)式,我们便得到如下著名的 **Jordan 定理**.

定理 4(Jordan 定理)　在前面记号下,变换 $T\in\mathscr{L}(\mathscr{S})$ 关于 $\mathscr{S}$ 的 T-Jordan 基的矩阵表示呈如下分块对角矩阵形式:

$$\boldsymbol{J}=\mathrm{diag}[J(\lambda_1),J(\lambda_2),\cdots,J(\lambda_s)],\tag{38}$$

这里,对 $i=1,\cdots,s$,

$$J(\lambda_i)=\mathrm{diag}[\underbrace{J_i^{(m_i)},\cdots,J_i^{(m_i)}}_{t_{m_i}^{(i)}\text{次}},\underbrace{J_i^{(m_i-1)},\cdots,J_i^{(m_i-1)}}_{t_{m_i-1}^{(i)}\text{次}},\cdots,\underbrace{J_i^{(1)},\cdots,J_i^{(1)}}_{t_1^{i}\text{次}}],\tag{39}$$

其中,$\sum_{k=1}^{m_i}t_k^{(i)}=t^{(i)}=\dim(\mathrm{Ker}(T-\lambda_iI))$,诸数 $t_k^{(i)}$ 由 T 唯一确定(如对某 $k<m_i$ 有 $t_k^{(i)}=0$,则(39)式右端应删掉对应的矩阵).

我们指出,由于(38)与(39)两式给出的 J 的分解中每一个 Jordan 块不可能相似于块数多于 1 的分块对角矩阵,在这种意义下,定理 4 给出变换 T 的最简单的矩阵表示 J. 通常,这样的 J 称为变换 $T\in\mathscr{L}(\mathscr{S})$ 的 **Jordan 标准型**或**正规形式**. 诸 $\lambda I-J_i$(这里 J_i 取遍(39)式右端的所有 Jordan 块)的初等因子称为 **$\lambda I-T$ 的初等因子**.

推论 5　任意一个矩阵 $A\in\boldsymbol{M}_n(\mathbb{C})$ 相似于一个形如(38)式的矩阵,且若不计

(38)与(39)两式右端诸矩阵的前后次序，此标准型矩阵由 A 唯一确定.

矩阵 $A\in \boldsymbol{M}_n(\mathbb{C})$ 的 Jordan 标准型(38)中，A 的特征值 λ_i 的代数重数等于矩阵 $J(\lambda_i)$ 的阶数或 $\dim(\mathrm{Ker}(A-\lambda_i I)^{m_i})$，这里，$m_i=\mathrm{index}\lambda_i$，$i=1,\cdots,s$. A 的特征值 λ_i 的几何重数等于 $J(\lambda_i)$ 中 Jordan 块的个数 $t^{(i)}=\dim(\mathrm{Ker}(A-\lambda_i I))$. 因此，任意 $\lambda\in\sigma(A)$ 的几何重数不会超过它的代数重数. 并且，当且仅当 $A\in \boldsymbol{M}_n(\mathbb{C})$ 为简单矩阵时，A 的每一个特征值的代数重数等于它的几何重数. 这表明简单矩阵没有级数大于 1 的广义特征向量.

例 6 确定下列 4×4 矩阵的 Jordan 正规形式：

$$A=\begin{bmatrix}2 & -1 & 0 & 1\\ 0 & 3 & -1 & 0\\ 0 & 1 & 1 & 0\\ 0 & -1 & 0 & 3\end{bmatrix}.$$

它的特征多项式 $\det(\lambda I-A)=(\lambda-2)^3(\lambda-3)$，因而 $\sigma(A)=\{2,3\}$，特征值 2 的代数重数为 3，其几何重数为 $\dim(\mathrm{Ker}(2I-A))=2$，因为 $\mathrm{rank}(2I-A)=2$. 特征值 3 的代数与几何重数均为 1. 于是，A 的 Jordan 正规形式为两个对应于特征值 2 的 Jordan 块与一个对应特征值 3 的一阶 Jordan 块之直和. 不计 Jordan 块的次序，它们的阶数只有一种可能：2，1，1.（实际上，为了确定 Jordan 块的阶数，我们由计算知道，$\mathrm{rank}(2I-A)^2=\mathrm{rank}(2I-A)^3=1$，因而 A 的最小多项式为 $m(\lambda)=(\lambda-2)^2(\lambda-3)$. 这表明 A 的三个 Jordan 块的阶数分别为 2，1，1）因此有

$$J=\mathrm{diag}\left[\begin{bmatrix}2 & 1\\ 0 & 2\end{bmatrix},[2],[3]\right].$$

经过计算顺便地得到，$\mathrm{Ker}(A-2I)$ 中有 A 的两个线性无关的特征向量 $\boldsymbol{\alpha}_1=(-1,-1,-1,-1)^{\mathrm{T}}$ 与 $\boldsymbol{\alpha}_2=(1,0,0,0)^{\mathrm{T}}$；$\mathrm{Ker}(A-2I)^2$ 中有基 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3=(0,1,2,0)^{\mathrm{T}}$，且 $(A-2I)\boldsymbol{\alpha}_3=\boldsymbol{\alpha}_1$；$\mathrm{Ker}(A-3I)$ 中有 A 的特征向量 $\boldsymbol{\alpha}_4=(1,0,0,1)^{\mathrm{T}}$. 它们组成 A 的 Jordan 基，其中，$\boldsymbol{\alpha}_1$ 与 $\boldsymbol{\alpha}_2=$ 为对应于特征值 2 的 A 的特征向量，$\boldsymbol{\alpha}_3$ 为 A 的(对应于特征值 2)二级广义特征向量，$\boldsymbol{\alpha}_4$ 为 A 的对应特征值 3 的特征向量. 令 $Q=[\boldsymbol{\alpha}_1\ \boldsymbol{\alpha}_3\quad \boldsymbol{\alpha}_2\quad \boldsymbol{\alpha}_4]$，则有 $A=QJQ^{-1}$. □

习题 1.1.2

1. 证明(10)式中两个等式成立.

2. 证明定理 1.

3. 试证：$T\in\mathscr{L}(\mathscr{S})$ 的最小多项式 $m(\lambda)$ 的不同零点组成 T 的谱 $\sigma(T)$.

4. 设 $T\in\mathscr{L}(\mathscr{S})$. 试证：对应于 $\lambda\in\sigma(T)$ 的长度为 r 的 Jordan 链中各元素线性无关.

5. 设 $T\in\mathscr{L}(\mathscr{S})$，$\mathscr{S}_i=\mathrm{Ker}(T-\lambda_i I)^{m_i}$，$\lambda_i\in\sigma(T)$，$m_i=\mathrm{index}\lambda_i$. 试证：若不计和项次序，$\mathscr{S}_i$ 有唯一分解(35).

6. 考虑下列矩阵

$$A=\begin{bmatrix}6&2&2\\-2&2&0\\0&0&2\end{bmatrix},\quad B=\begin{bmatrix}1&0&-1&1&0\\-4&1&-3&2&1\\-2&-1&0&1&1\\-3&-1&-3&4&1\\-8&-2&-7&5&4\end{bmatrix}$$

的 Jordan 正规形式,并求出 A 与 B 各特征值的代数与几何重数以及有关的广义特征向量.

7. 试验证:当 $\alpha_1\neq 0$ 时,$(\lambda-\alpha_0)^r$ 为矩阵 $\lambda I-A$ 的唯一初等因子,其中

$$A=\begin{bmatrix}\alpha_0&\alpha_1&\cdots&\alpha_{r-1}\\0&\alpha_0&&\vdots\\&\ddots&\ddots&\ddots\\\vdots&&&\alpha_1\\0&\cdots&0&\alpha_0\end{bmatrix}\in \boldsymbol{M}_r(\mathbf{C}).$$

8. 设 $A\in\boldsymbol{M}_r(\mathbf{C})$ 如题 7,但 $a_1=a_2=\cdots=a_{l-1}=0,a_l\neq 0(1\leqslant l\leqslant r-1)$. 试证:

(1) $l_s\equiv\dim(\mathrm{Ker}(a_0I-A)^s)=\begin{cases}ls, & \text{如果 } ls\leqslant r,\\ r, & \text{如果 } ls>r.\end{cases}$

(2) $l_1=l,l_2=2l,\cdots,l_p=pl,l_{p+i}=r,i\geqslant 1$,这里,$p$ 为 r/l 的整数部分.

(3) $\lambda I-A$ 的初等因子为

$$\underbrace{(\lambda-a_0)^p,\cdots,(\lambda-a_0)^p}_{l-q\text{次}},\ \underbrace{(\lambda-a_0)^{p+1},\cdots,(\lambda-a_0)^{p+1}}_{q\text{次}},$$

其中,整数 p 与 q 满足条件 $r=pl+q(0\leqslant q\leqslant l-1,p>0)$.

9. 试证:若 $J\in\boldsymbol{M}_r(\mathbf{C})$ 为对应于特征值 λ_0 的 Jordan 块,则对任意复系数多项式 $p(\lambda)$ 有

$$p(J)=\begin{bmatrix}p(\lambda_0)&p'(\lambda_0)/1!&\cdots&p^{(r-1)}(\lambda_0)/(r-1)!\\0&p(\lambda_0)&\ddots&\vdots\\\vdots&\vdots&\ddots&p'(\lambda_0)/1!\\0&\cdots&0&p(\lambda_0)\end{bmatrix}.$$

根据这个结果证明:若 $|\lambda_0|<1$,则当 $k\to\infty$ 时,J^k 的所有元素都趋于零.

10. 试证:若 $A\in\boldsymbol{M}_n(\mathbf{C})$,则 $A=A^{-1}$ 当且仅当存在 $m\leqslant n$ 与非奇异矩阵 S 使得

$$A=S\begin{bmatrix}I_m&O\\O&-I_{n-m}\end{bmatrix}S^{-1}.$$

11. 试证:n 阶友阵(见习题 1.1.1 题 13)是**非减阶的**,即它的每个不同特征值 λ_j 对应且仅对应着一个 $m_j(=\mathrm{index}\lambda_j)$ 阶的 Jordan 块,或等价地,它的最小多项式的次数为 n.

1.2 对称矩阵与 Hermite 矩阵,酉空间上的线性变换

本节讨论有广泛应用的对称矩阵与 Hermite 矩阵的基本性质. 更一般地,我

们还讨论自酉空间到另一个酉空间的线性变换的一些性质，证明酉空间上的线性变换有伴随变换. 应用伴随变换可以定义与研究正规变换与正规矩阵、自伴变换与 Hermite 矩阵，以及酉变换与酉矩阵.

1.2.1 正规变换与正规矩阵

设 $\mathscr{S}$ 为域 $\boldsymbol{F}$ 上的有限维线性空间. 当 $\boldsymbol{F}=\mathbb{C}$ 时，带有从 $\mathscr{S}\times\mathscr{S}$ 到 $\mathbb{C}$ 的内积 $\langle,\rangle$ 的复线性空间 $\mathscr{S}$ 叫作**酉空间**(unitary space)；当 $\boldsymbol{F}=\mathbb{R}$ 时，带有从 $\mathscr{S}\times\mathscr{S}$ 到 $\mathbb{R}$ 的内积 $\langle,\rangle$ 的实线性空间 $\mathscr{S}$ 叫作**欧氏空间**(Euclidean space). 当有必要区分不同内积时，将用 $\mathscr{S}\langle,\rangle$ 表示有内积 $\langle,\rangle$ 的由 $\mathscr{S}$ 生成的酉空间或欧氏空间. 当考虑 $\mathbb{C}^n$ 或 $\mathbb{R}^n$ 为酉空间或欧氏空间时，如无另外声明，它的内积总是指所谓**标准内积**：

$$\langle \boldsymbol{x},\boldsymbol{y}\rangle = \boldsymbol{y}^*\boldsymbol{x}, \qquad \forall\, \boldsymbol{x},\boldsymbol{y}\in\mathbb{C}^n,$$
$$\langle \boldsymbol{x},\boldsymbol{y}\rangle = \boldsymbol{y}^{\mathrm{T}}\boldsymbol{x}, \qquad \forall\, \boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n. \tag{1}$$

若 $\mathscr{U}=\mathscr{S}\langle,\rangle$ 为酉空间，则可以用 $\langle \boldsymbol{x},\boldsymbol{x}\rangle^{1/2}$ 定义 $\boldsymbol{x}\in\mathscr{U}$ 的**长度**或**范数**，记为 $\|\boldsymbol{x}\|$；用 $\langle \boldsymbol{x},\boldsymbol{y}\rangle=0$ 定义 $\boldsymbol{x},\boldsymbol{y}\in\mathscr{U}$ **正交**. $\mathscr{U}$ 中非零元素的集合叫作**正交系**，假如它的任意两个不同元素正交；若正交系内每一个元素的范数均为 1，则此正交系叫作**标准的**. 显然，$\mathscr{U}$ 中任一正交系的元素线性无关. 另一方面，$\mathscr{U}$ 中任一线性无关元素的集合 $\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_p\}$ 可以经过如下的所谓 **Gram-Schmidt 正交化过程**变为 $\mathscr{U}$ 的一个正交系 $\{\boldsymbol{y}_1,\boldsymbol{y}_2,\cdots,\boldsymbol{y}_p\}$：

$$\boldsymbol{y}_1=\boldsymbol{x}_1,\quad \boldsymbol{y}_r=\boldsymbol{x}_r-\sum_{j=1}^{r-1}\frac{\langle \boldsymbol{y}_j,\boldsymbol{x}_r\rangle}{\langle \boldsymbol{y}_j,\boldsymbol{y}_j\rangle}\boldsymbol{y}_j,\qquad r=2,\cdots,p. \tag{2}$$

于是，维数非零的每一个有限维酉空间有一个正交系的基，进而有一个标准正交系的基，简称**标准正交基**.

若 $\mathscr{S}_1$ 为酉空间 $\mathscr{U}=\mathscr{S}\langle,\rangle$ 的线性子空间，则

$$\mathscr{S}_2 \equiv \{\boldsymbol{x}\in\mathscr{U}: \langle \boldsymbol{x},\boldsymbol{y}\rangle=0, \forall\, \boldsymbol{y}\in\mathscr{S}_1\} \tag{3}$$

为 $\mathscr{U}$ 的子空间，称之为 $\mathscr{S}_1$ 在 $\mathscr{U}$ 中的**正交补**，记为 $\mathscr{S}_2=\mathscr{S}_1^{\perp}$. 容易看出，$(\mathscr{S}_1^{\perp})^{\perp}=\mathscr{S}_2^{\perp}=\mathscr{S}_1$，因而 $\mathscr{S}_1$ 与 $\mathscr{S}_2$ 是 $\mathscr{U}$ 中互为正交补的子空间，这时 $\mathscr{U}$ 表示为 $\mathscr{S}_1$ 与 $\mathscr{S}_2$ 的**正交和**，记为

$$\mathscr{U}=\mathscr{S}_1\oplus\mathscr{S}_2. \tag{4}$$

类似地，可以给出 $\sum\limits_{i=1}^{p}\oplus\mathscr{S}_i=\mathscr{S}_1\oplus\mathscr{S}_2\oplus\cdots\oplus\mathscr{S}_p$ 的定义.

显然，1.1.2 小节的结果也适用于线性变换 $T\in\mathscr{L}(\mathscr{U},\mathscr{V})$，这里，$\mathscr{U}=\mathscr{S}\langle,\rangle_1$ 与 $\mathscr{V}=\mathscr{T}\langle,\rangle_2$. 对这样的线性变换 T，我们可以定义它的对偶变换，具体如下：

$$\langle T(\boldsymbol{x}),\boldsymbol{y}\rangle_2=\langle \boldsymbol{x},T^*(\boldsymbol{y})\rangle_1,\ \forall\,\boldsymbol{x}\in\mathscr{U},\boldsymbol{y}\in\mathscr{V}. \tag{5}$$

上式中 T^* 称为 T 的**伴随**(adjoint). 容易证明，$T^*\in\mathscr{L}(\mathscr{V},\mathscr{U})$ 是唯一的，且有

$$I^*=I,\ O^*=O,(T^*)^*=T,(\alpha S+\beta T)^*=\bar{\alpha}S^*+\bar{\beta}T^*,$$

$$(TS)^* = S^* T^*, \quad (T^*)^{-1} = (T^{-1})^* \text{(若 } T \text{ 为可逆的)}.$$

并且,若 $A \in \boldsymbol{M}_{m,n}(\boldsymbol{F})$ 为 $T \in \mathscr{L}(\mathscr{U},\mathscr{V})$ 关于标准正交基对 $(\mathscr{E},\mathscr{G})$ 的矩阵表示,则 $A^* \in \boldsymbol{M}_{n,m}(\boldsymbol{F})$ 便为 $T^* \in \mathscr{L}(\mathscr{V},\mathscr{U})$ 关于基对 $(\mathscr{G},\mathscr{E})$ 的矩阵表示,即有

$$_{\mathscr{E}}[T^*]_{\mathscr{G}} = {}_{\mathscr{G}}[T]_{\mathscr{E}}^*. \tag{6}$$

特别地,当 $T \in \mathscr{L}(\mathscr{U})$ 时,且 $\mathscr{E}$ 与 $\mathscr{E}'$ 为 $\mathscr{U}$ 的一对双正交基(即 $\mathscr{E}=\{\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n\}$ 与 $\mathscr{E}'=\{\boldsymbol{y}_1,\cdots,\boldsymbol{y}_n\}$ 为满足条件 $\langle \boldsymbol{x}_i,\boldsymbol{y}_j\rangle=\delta_{ij}, 1\leqslant i,j\leqslant n$,的 $\mathscr{U}$ 的一对正交基),我们有

$$[T^*]_{\mathscr{E}'} = [T]_{\mathscr{E}}^*. \tag{7}$$

如取 $\mathscr{E}=\mathscr{E}'$ 为 $\mathscr{U}$ 的标准正交基,则 T 与 T^* 关于同一标准正交基的表示满足

$$[T^*]_{\mathscr{E}} = [T]_{\mathscr{E}}^*. \tag{8}$$

对任意 $T \in \mathscr{L}(\mathscr{U},\mathscr{V})$,由 1.1.2 小节(10)式与这里的(6)式得出,$\mathrm{rank}T=\mathrm{rank}T^*$,且若 $\mathscr{U}=\mathscr{V}$,则还有 $\mathrm{nul}T=\mathrm{nul}T^*$.

设 $T \in \mathscr{L}(\mathscr{U},\mathscr{V})$ 且 T^* 为 T 的伴随变换,则酉空间 $\mathscr{U}$ 与 $\mathscr{V}$ 分别有如下的正交分解:

$$\mathscr{U} = \mathrm{Ker}T \oplus \mathrm{Im}T^*, \qquad \mathscr{V} = \mathrm{Ker}T^* \oplus \mathrm{Im}T, \tag{9}$$

因为 $\langle \boldsymbol{x},T(\boldsymbol{y})\rangle_2=0, \forall\, \boldsymbol{y}\in\mathscr{U}$ 蕴涵 $\boldsymbol{x}\in\mathrm{Ker}T^*$. 因此,

$$\mathrm{rank}T + \mathrm{nul}T^* = \dim\mathscr{V}. \tag{10}$$

并且,对于 $T \in \mathscr{L}(\mathscr{U},\mathscr{V})$ 来说,或者方程 $T(\boldsymbol{x})=\boldsymbol{b}$ 可解,$\forall\,\boldsymbol{b}\in\mathscr{V}$,或者方程 $T^*(\boldsymbol{y})=\boldsymbol{0}$ 有非零解. 事实上,若 $T(\boldsymbol{x})=\boldsymbol{b}$ 对任意 $\boldsymbol{b}\in\mathscr{V}$ 可解,那么,$\mathrm{Im}T=\mathscr{V}$,因而由(9)式推出 $\mathrm{Ker}T^*=\{\boldsymbol{0}\}$,即 $T^*(\boldsymbol{y})=\boldsymbol{0}$ 只有零解. 反之,类似地也可证明 $T^*(\boldsymbol{y})=\boldsymbol{0}$ 有非零解蕴涵 $T(\boldsymbol{x})=\boldsymbol{b}$ 对某 $\boldsymbol{b}\in\mathscr{V}$ 无解,因为 $\mathrm{Im}T \underset{\neq}{\subset} \mathscr{V}$.

特殊地,若 $A\in\boldsymbol{M}_{m,n}(\mathbb{C})$(视为 $\mathscr{L}(\mathbb{C}^n,\mathbb{C}^m)$ 的变换),则从(9)式推出

$$\mathbb{C}^n = \mathrm{Ker}A \oplus \mathrm{Im}A^* \text{ 与 } \mathbb{C}^m = \mathrm{Ker}A^* \oplus \mathrm{Im}A. \tag{9'}$$

假定 $\mathscr{U}$ 为 n 维酉空间,$T\in\mathscr{L}(\mathscr{U})$ 叫作**正规变换**,假如 $\mathscr{U}$ 存在由 T 的特征向量组成的标准正交基,简称**标准正交特征基**. 同样地,矩阵 $A\in\boldsymbol{M}_n(\mathbb{C})$ 叫作**正规的**,假如将 A 看成为 $\mathscr{L}(\mathbb{C}^n)$ 中变换 T_A(见 1.1.2 小节(16)式),T_A 为正规的. 一个变换 $T\in\mathscr{L}(\mathscr{U})$ 为正规的当且仅当 T 关于 $\mathscr{U}$ 的任意标准正交基 $\mathscr{E}$ 的表示 $[T]_{\mathscr{E}}$ 为正规矩阵. 这是因为 $\boldsymbol{y}=T(\boldsymbol{x}), \boldsymbol{x}\in\mathscr{U}$ 等价于 $[\boldsymbol{y}]_{\mathscr{E}}=[T]_{\mathscr{E}}[\boldsymbol{x}]_{\mathscr{E}}$,这里,$\mathscr{E}$ 为 $\mathscr{U}$ 的任一标准正交基,并且 $\langle \boldsymbol{x},\boldsymbol{z}\rangle=[\boldsymbol{z}]_{\mathscr{E}}^*[\boldsymbol{x}]_{\mathscr{E}}=\langle [\boldsymbol{x}]_{\mathscr{E}},[\boldsymbol{z}]_{\mathscr{E}}\rangle, \forall\,\boldsymbol{x},\boldsymbol{z}\in\mathscr{U}$,这里 $\mathscr{E}$ 为 $\mathscr{U}$ 的任一标准正交基.

当 $\mathscr{U}$ 为酉空间时,今后用 $\mathscr{A}_T^0$ 专指 $T\in\mathscr{L}(\mathscr{U})$ 关于 $\mathscr{U}$ 的任意标准正交基 $\mathscr{E}$ 的矩阵表示 $[T]_{\mathscr{E}}$ 的集合. 这时,$\mathscr{A}_T^0$ 中任意两个矩阵 A 与 B 满足 $B=UAU^{-1}$,不过其中的 U 是从 $\mathscr{U}$ 的一个标准正交基到另一个标准正交基的特殊的转移矩阵,因而 U 的列向量组成 $\mathbb{C}^n$ 的标准正交基,即有 $U^*U=I$,这样的矩阵 U 通常叫作**酉矩阵**. 因此,$B=UAU^*$,即 A 与 B 为**酉相似的**. 这表明,$\mathscr{A}_T^0$ 恰为 $T\in\mathscr{L}(\mathscr{U})$ 关于 $\mathscr{U}$ 的标准正

交基的矩阵表示的等价类 $A,B\in\mathscr{A}_T^0$ 等价于 A 与 B 的酉相似性.

当 $T\in\mathscr{L}(\mathscr{U})$ 为正规变换时,我们自然希望得到 $\mathscr{A}_T^0$ 中正规矩阵表示的最简单形式,即酉相似意义下的标准型.下面定理讨论这个问题,它指出正规矩阵酉相似标准型为对角矩阵.

定理 1 设 $A\in\boldsymbol{M}_n(\mathbb{C})$. 则 A 为正规矩阵当且仅当 A 酉相似于一个对角矩阵 D,而 D 的对角元素为 A 的 n 个特征值 $\lambda_1,\cdots,\lambda_n$,换言之,

$$A = UDU^*, \tag{11}$$

这里,U 为酉矩阵,$D=\mathrm{diag}[\lambda_1,\cdots,\lambda_n]$,$\lambda_i\in\sigma(A)$,$i=1,\cdots,n$.

证明 设 $A\in\boldsymbol{M}_n(\mathbb{C})$ 为正规矩阵. 按定义它有特征向量 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n$ 组成的酉空间 $\mathbb{C}^n$ 的标准正交基 $\mathscr{E}$. 令 $A\boldsymbol{x}_j=\lambda_j\boldsymbol{x}_j$,$\lambda_j\in\sigma(A)$,$1\leqslant j\leqslant n$. 则 $U\equiv[\boldsymbol{x}_1\cdots\ \boldsymbol{x}_n]\in\boldsymbol{M}_n(\mathbb{C})$ 为酉矩阵,因为 $U^*U=[\delta_{ij}]=I_n$,且有 $AU=UD$,$D=\mathrm{diag}[\lambda_1,\cdots,\lambda_n]$,因而(11)式成立. 反之,若(11)式成立,则 U 的每一个列向量为 A 的特征向量,再由 U 为酉矩阵的事实直接推出这组 A 的特征向量构成 $\mathbb{C}^n$ 的标准正交基,因而 $A\in\boldsymbol{M}_n(\mathbb{C})$ 是正规矩阵. □

$\boldsymbol{M}_n(\mathbb{R})$ 中酉矩阵称为**正交矩阵**,它满足条件 $U^{\mathrm{T}}U=I_n$. 若 $A,B\in\boldsymbol{M}_n(\mathbb{R})$ 满足 $B=UAU^{\mathrm{T}}$,这里,$U\in\boldsymbol{M}_n(\mathbb{R})$ 为正交矩阵,则 A 与 B 称为**正交相似的**. $\boldsymbol{M}_n(\mathbb{R})$ 中正规矩阵的正交相似标准型一般不是(11)式中的对角矩阵 D(见习题 1).

当 $A\in\boldsymbol{M}_n(\mathbb{C})$ 不是正规矩阵时,按定理 1,(11)式不成立,但有如下重要结论.

定理 2(Schur 三角化定理) 若 $A\in\boldsymbol{M}_n(\mathbb{C})$,则 A 酉相似于一个上三角形矩阵 B,而 B 的对角元素均为 A 的特征值,亦即

$$A = UBU^*, \tag{12}$$

这里,U 为酉矩阵,$B\in\boldsymbol{M}_n(\mathbb{C})$ 为上三角形矩阵,它的对角元素均为 A 的特征值. 若 $A\in\boldsymbol{M}_n(\mathbb{R})$ 且它的特征值均为实数,则(12)式中 U 可取为实正交矩阵,即 A 正交相似于一个上三角形矩阵 B.

有了上述两个定理容易得到下列结论.

推论 3 (1) 设 $A\in\boldsymbol{M}_n(\mathbb{C})$,则 A 为正规矩阵当且仅当

$$AA^* = A^*A. \tag{13}$$

(2) 设 $T\in\mathscr{L}(\mathscr{U})$,则 T 为正规变换当且仅当

$$TT^* = T^*T. \tag{14}$$

推论 4 若 $A\in\boldsymbol{M}_n(\mathbb{C})$ 为正规矩阵,则 $\boldsymbol{x}$ 为 A 的特征向量当且仅当它为 A^* 的特征向量,且 $A\boldsymbol{x}=\lambda\boldsymbol{x}$ 蕴涵 $A^*\boldsymbol{x}=\bar{\lambda}\boldsymbol{x}$. 此外,对应不同特征值的 A 的特征向量彼此正交.

按 1.1.1 小节中简单矩阵的定义与这里的定理 1,我们推出,任意正规矩阵 $A\in\boldsymbol{M}_n(\mathbb{C})$ 必定是简单矩阵. 因此,由 1.1.1 小节定理 7 可得出正规矩阵的谱定理.

定理 5(正规矩阵的谱定理)　设 $A\in \boldsymbol{M}_n(\mathbb{C})$ 为正规矩阵，它有特征值 $\lambda_1,\cdots,\lambda_n$. 则存在对应的标准正交特征向量 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n$ 使得 A 有如下的谱形式：

$$A=\sum_{j=1}^{n}\lambda_j G_j, \tag{15}$$

这里，$G_j=\boldsymbol{x}_j\boldsymbol{x}_j^*$ 为秩 1 的 n 阶 Hermite 矩阵，且 $G_iG_j=\delta_{ij}G_i$, $\forall\, i,j$.

$\mathscr{L}(\mathscr{U})$ 中正规变换的两种特殊情形，即自伴变换与酉变换在实用中起着重要的作用. 我们说，$T\in\mathscr{L}(\mathscr{U})$ 为**自伴变换**，假若 $T=T^*$；$T\in\mathscr{L}(\mathscr{U})$ 为**酉变换**，假若 $TT^*=T^*T=I$. 容易验证，$T\in\mathscr{L}(\mathscr{U})$ 为自伴(酉)变换当且仅当 $\mathscr{A}_T$ 中任意矩阵 A 为 Hermite(酉)矩阵.

Hermite 矩阵具有许多重要的性质. 首先，按定理 1，正规矩阵 $A\in\boldsymbol{M}_n(\mathbb{C})$ 为 Hermite 矩阵当且仅当它的谱属于 $\mathbb{R}$. 这时，(11)式中 $D=\mathrm{diag}[\lambda_1,\cdots,\lambda_n]$ 为实对角矩阵. 其次，从定理 5 经过更详细讨论可得到 Hermite 矩阵的谱定理.

定理 6(Hermite 矩阵的谱定理)　设 $A\in\boldsymbol{M}_n(\mathbb{C})$ 为 Hermite 矩阵. 则存在正整数 s, $1\leqslant s\leqslant n$, s 个幂等的非零 Hermite 矩阵 $E_1,\cdots,E_s$ 与 s 个数 $\lambda_1,\cdots,\lambda_s$ 使得

(1) $E_iE_j=O$, $\forall\, i\neq j$，即诸 E_j 彼此正交.

(2) $\lambda_1,\cdots,\lambda_s$ 彼此不同.

(3) $\displaystyle\sum_{j=1}^{s}E_j=I_n$.

(4) $AE_j=E_jA$, $j=1,\cdots,s$.

(5) $A=\displaystyle\sum_{j=1}^{s}\lambda_jE_j$(**Hermite 矩阵的谱形式**).

式中诸 λ_j 与 E_j 由上述条件(1)～(5)唯一确定. 此外，

(6) $\lambda_1,\cdots,\lambda_s$ 恰为 A 的所有不同的特征值，因而它们均为实数.

(7) $\dim(\mathrm{Im}E_j)=\mathrm{rank}E_j$ 恰等于 $\lambda_j\in\sigma(A)$ 的代数重数，$i=1,\cdots,s$.

(8) $BA=AB$ 当且仅当 $BE_j=E_jB$, $j=1,\cdots,s$.

证明　设 $\sigma(A)=\{\lambda_1,\cdots,\lambda_s\}$. 将(15)式中对应于同一特征值的 $G_j=\boldsymbol{x}_j\boldsymbol{x}_j^*$ 合并便得 $E_1,\cdots,E_s$. 这时，条件(1),(2),(4)与(5)由定理 5 直接可得. 为证条件(3)，可令 $\displaystyle\sum_{j=1}^{s}E_j=E$. 由条件(1)，$E^2=E$ 且 $E=E^*$. 若 $E\neq I_n$，则 A 在 $\mathrm{Im}(I_n-E)$ 上的限制(视为线性变换)没有特征值，这是不可能的. 现在证明条件(5)表达式中诸 λ_j 与 E_j 是由条件(1)～(5)唯一确定的. 任取 $\boldsymbol{x}\in\mathrm{Im}E_j$，则有 $E_j\boldsymbol{x}=\boldsymbol{x}$ 与 $E_i\boldsymbol{x}=\boldsymbol{0}$, $\forall\, i\neq j$，因而

$$A\boldsymbol{x}=\sum_{i=1}^{s}\lambda_iE_i\boldsymbol{x}=\lambda_jE_j\boldsymbol{x}=\lambda_j\boldsymbol{x}.$$

于是，$\lambda_j\in\sigma(A)$. 反之，若 $\lambda\in\sigma(A)$，且有 $\boldsymbol{x}\neq\boldsymbol{0}$ 使得 $A\boldsymbol{x}=\lambda\boldsymbol{x}$，记 $\boldsymbol{x}_j=E_j\boldsymbol{x}$，则有

$$Ax = \lambda x = \lambda \sum_{j=1}^{s} E_j x = \lambda \sum_{j=1}^{s} x_j$$

与

$$Ax = A\sum_{j=1}^{s} E_j x = \Big(\sum_{j=1}^{s} \lambda_i E_i\Big) \sum_{j=1}^{s} E_j x = \sum_{j=1}^{s} \lambda_j x_j.$$

因此，$\sum_j (\lambda - \lambda_j) x_j = \mathbf{0}$. 因为 $x_1, \cdots, x_s$ 两两正交，它们中间非零向量组成线性无关集合，所以我们有 $x_j = \mathbf{0}$ 或者 $\lambda = \lambda_j$. 但由于 $x \neq \mathbf{0}$，从条件(3)便知有某个 $x_j \neq \mathbf{0}$，因而 λ 一定为某个 λ_j. 上述证明表明，A 的表达式 $\sum_j \lambda_j E_j$ 中 $\lambda_1, \cdots, \lambda_s$ 恰为 A 的所有不同的特征值. 因此结论(b)成立. 接着证明诸 E_j 的唯一性. 由于 $E_1, \cdots, E_s$ 为彼此正交的幂等非零的 Hermite 矩阵，故有 $A^k = \sum_j \lambda_j^k E_j, k = 1, 2, \cdots$，因而对任意复系数多项式 $p(\lambda)$，我们有 $p(A) = \sum_j p(\lambda_j) E_j$. 现取特殊多项式

$$\widetilde{p}_i(\lambda) = \prod_{j \neq i} \frac{\lambda - \lambda_j}{\lambda_i - \lambda_j}, i = 1, \cdots, s. \tag{16}$$

它满足 $\widetilde{p}_i(\lambda_i) = 1$ 与 $\widetilde{p}_i(\lambda_j) = 0, \forall j \neq i$. 对此 $\widetilde{p}_i(\lambda)$，我们有 $\widetilde{p}_i(A) = E_i$. 这表明 E_i 为 A 的多项式，而此多项式 $\widetilde{p}_i(\lambda)$ 由 A 的 s 个不同特征值唯一确定. 按此我们也顺便地证明了(7)与(8). □

Hermite 矩阵 $A \in \boldsymbol{M}_n(\mathbb{C})$ 的实特征值可按大小排列为 $\lambda_1 \geqslant \lambda_2 \cdots \geqslant \lambda_n$(计入重根). 后面将证明，这些特征值可以用 A 的 **Rayleigh 商**

$$R_A(x) = \langle Ax, x\rangle / \langle x, x\rangle, x \neq \mathbf{0} \tag{17}$$

来表示. 因为 $\langle Ax, x\rangle = \langle x, A^* x\rangle = \langle x, Ax\rangle = \overline{\langle Ax, x\rangle}$，所以对所有 $\mathbf{0} \neq x \in \mathbb{C}^n$，$R_A(x)$ 为实值. 显然

$$\min_{\langle x,x\rangle = 1} \langle Ax, x\rangle = \min_{x \neq \mathbf{0}} R_A(x), \tag{18}$$

$$\max_{\langle x,x\rangle = 1} \langle Ax, x\rangle = \max_{x \neq \mathbf{0}} R_A(x). \tag{19}$$

下面定理指出，(18)式等于 λ_n，(19)式等于 λ_1；A 的所有特征值可以用 Rayleigh 商的极小-极大值或极大-极小值表示.

定理 7(Courant-Fischer) 设 $A \in \boldsymbol{M}_n(\mathbb{C})$ 为 Hermite 矩阵，它有特征值 $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$. 又设 $\mathscr{S}_m$ 表示 $\mathbb{C}^n$ 的任意 $n-m+1$ 维子空间. 则

$$\lambda_m = \min_{\mathscr{S}_m} \max_{0 \neq x \in \mathscr{S}_m} R_A(x), m = 1, \cdots, n, \tag{20}$$

$$\lambda_{n-m+1} = \max_{\mathscr{S}_m} \min_{0 \neq x \in \mathscr{S}_m} R_A(x), m = 1, \cdots, n. \tag{21}$$

证明 设 $x_1, x_2, \cdots, x_n$ 为对应于特征值 $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$ 的 A 的标准正交特征基，且令 $\hat{\mathscr{S}}_m = \mathrm{span}\{x_1, \cdots, x_m\}, m = 1, \cdots, n$. 对任意给定 m，由于 $\dim \mathscr{S}_m + \dim \hat{\mathscr{S}}_m = (n-m+1) + m = n+1 > n$，我们有 $\mathbf{0} \neq x_0 \in \mathbb{C}^n$ 同时属于 $\mathscr{S}_m$ 与 $\hat{\mathscr{S}}_m$，且有 $\| x_0 \| = 1$.

$x_0 \in \hat{\mathscr{S}}_m$ 蕴涵 $x_0 = \sum_{k=1}^{m} \alpha_k x_k$，因而

$$\langle Ax_0, x_0 \rangle = \sum_{k=1}^{m} \lambda_k \mid \alpha_k \mid^2 \geqslant \lambda_m \sum_{k=1}^{m} \mid \alpha_k \mid^2 = \lambda_m.$$

于是 $\max\limits_{0 \neq x \in \mathscr{S}_m} R_A(x) \geqslant \lambda_m$. 另一方面，考虑特殊的 $n-m+1$ 维子空间 $\tilde{\mathscr{S}}_m = \operatorname{span}\{x_m, x_{m+1}, \cdots, x_n\}$，对此 $\tilde{\mathscr{S}}_m$ 中任意的 $x = \sum_{k=m}^{n} \alpha_k x_k$，我们有

$$\langle Ax, x \rangle = \sum_{k=m}^{n} \lambda_k \mid \alpha_k \mid^2 \leqslant \lambda_m \sum_{k=m}^{n} \mid \alpha_k \mid^2 = \lambda_m \langle x, x \rangle,$$

这表明 $\max\limits_{0 \neq x \in \tilde{\mathscr{S}}_m} R_A(x) \leqslant \lambda_m$. 因此，当 $\mathscr{S}_m$ 取遍 $\mathbb{C}^n$ 的所有的 $n-m+1$ 维子空间时，$\max\limits_{0 \neq x \in \mathscr{S}_m} R_A(x)$ 总是 $\geqslant \lambda_m$ 且有某 $\mathscr{S}_m$ 使它 $\leqslant \lambda_m$，即有(20)式成立. 类似可证(21)式成立. □

我们指出，当 $m=1$ 时，$\mathscr{S}_1 = \mathbb{C}^n$，(20)与(21)式分别退化到 $\max\limits_{x \neq 0} R_A(x) = \lambda_1$ 与 $\min\limits_{x \neq 0} R_A(x) = \lambda_n$. 取 $x = e_k = (0, \cdots, 0, 1, 0, \cdots, 0)^T \in \mathbb{C}^n$，则有 $\lambda_n \leqslant a_{kk} \leqslant \lambda_1$，$k = 1, \cdots, n$，因而有

$$\lambda_1 \geqslant \max_k a_{kk},$$
$$\lambda_n \leqslant \min_k a_{kk}. \tag{22}$$

定理 7 的一个有用推论是如下的关于 Hermite 矩阵特征值的分离结果.

定理 8　设 $A \in \boldsymbol{M}_n(\mathbb{C})$ 为 Hermite 矩阵，它有特征值 $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$. 又设 B 为 A 的任意一个 k 阶主子阵，它有特征值 $\mu_1 \geqslant \mu_2 \geqslant \cdots \geqslant \mu_k$. 则

$$\lambda_m \geqslant \mu_m \geqslant \lambda_{n+m-k}, m = 1, \cdots, k. \tag{23}$$

上述定理的证明留给读者(见习题 2).

对于酉矩阵，首先从定理 1 看出，若 $U \in \boldsymbol{M}_n(\mathbb{C})$ 为正规矩阵，则当且仅当 $\sigma(U)$ 包含在复平面单位圆周上时，U 是酉矩阵.（实际上从定理 5 也可以推得这个结果）其次，从条件 $U^*U = I_n$ 看出，酉矩阵 U 的 n 个列向量组成 $\mathbb{C}^n$ 的一个标准正交基. 反之，若一个 n 阶复矩阵 U 的 n 个列向量组成 $\mathbb{C}^n$ 的一个标准正交基，则 U 为酉矩阵. 另外，同阶的有限多个酉矩阵之积仍为酉矩阵；任意酉矩阵 U 有逆 $U^* = U^{-1}$，且逆也是酉矩阵. 因此，在矩阵乘法运算之下，所有 n 阶酉矩阵组成一个群，该群的乘法单位元即是 n 阶单位矩阵 I_n.

酉矩阵另外重要特征可以用 $\mathbb{C}^n$ 的标准内积与欧氏长度(或范数)来描述. 这是下面基本定理的一部分内容.

定理 9　设 $U \in \boldsymbol{M}_n(\mathbb{C})$. 则下面诸命题彼此等价：

(1) U 为酉矩阵.

(2) $\| Ux \| = \| x \|$，$\forall x \in \mathbb{C}^n$.

(3) $\langle Ux,Uy\rangle=\langle x,y\rangle$, $\forall x,y\in\mathbb{C}^n$.

(4) 若 $u_1,\cdots,u_n$ 为$\mathbb{C}^n$ 的一个标准正交基,则 $Uu_1,\cdots,Uu_n$ 也是$\mathbb{C}^n$ 的一个标准正交基.

(5) 存在$\mathbb{C}^n$ 的一个标准正交基 $f_1,\cdots,f_n$,使得 $Uf_1,\cdots,Uf_n$ 也是$\mathbb{C}^n$ 的一个标准正交基.

证明 (1)⇒(2):设 U 为酉矩阵且 $x\in\mathbb{C}^n$. 则$\langle Ux,Ux\rangle=\langle x,U^*Ux\rangle=\langle x,x\rangle$,因而 $\|Ux\|=\|x\|$.

(2)⇒(3):对 $x,y\in\mathbb{C}^n$,应用恒等式

$$\langle x,y\rangle=\frac{1}{4}[\|x+y\|^2-\|x-y\|^2+\mathrm{i}\|x+\mathrm{i}y\|^2-\mathrm{i}\|x-\mathrm{i}y\|^2],$$

则上式右端按(2)当 x 与 y 分别换为 Ux 与 Uy 时,其值不变,因而$\langle Ux,Uy\rangle=\langle x,y\rangle$, $\forall x,y\in\mathbb{C}^n$.

(3) ⇒(4):设 $u_1,\cdots,u_n$ 为$\mathbb{C}^n$ 的一个标准正交基,且令 $v_i=Uu_i,i=1,\cdots,n$. 按(3)我们有

$$\langle v_i,v_j\rangle=\langle Uu_i,Uu_j\rangle=\langle u_i,u_j\rangle=\delta_{ij},$$

因而 $Uu_1,\cdots,Uu_n$ 为$\mathbb{C}^n$ 的一个标准正交基.

(4) ⇒(5):这是显然的,因为$\mathbb{C}^n$ 有一个标准正交基.

(5) ⇒(1):设 $f_1,\cdots,f_n$ 为$\mathbb{C}^n$ 的一个标准正交基,且 $Uf_1,\cdots,Uf_n$ 也是$\mathbb{C}^n$ 的一个标准正交基. 则有$\langle Uf_i,Uf_j\rangle=\delta_{ij}=\langle f_i,f_j\rangle$,因而$\langle f_i,U^*Uf_j\rangle=\langle f_i,f_j\rangle$或即

$$\langle f_i,(U^*U-I)f_j\rangle=0,\ \forall i,j.$$

由此推出$\langle x,(U^*U-I)f_j\rangle=0$, $\forall x\in\mathbb{C}^n,j=1,\cdots,n$. 此式蕴涵$(U^*U-I)f_j=\mathbf{0}$, $j=1,\cdots,n$,从而$(U^*U-I)x=\mathbf{0}$, $\forall x\in\mathbb{C}^n$,因为 $f_1,\cdots,f_n$ 为$\mathbb{C}^n$ 的基. 因此,$U^*U=I$,即 $U\in M_n(\mathbb{C})$为酉矩阵. □

我们指出,前面有关酉矩阵的结果对实正交矩阵也成立. 因此,实正交矩阵的特征值之积(即实正交矩阵的行列式)等于 1 或-1,而酉矩阵行列式的模等于 1.

下面例子提供实正交矩阵与酉矩阵的重要实例.

例 10 二阶矩阵

$$A_\alpha=\begin{bmatrix}\cos\alpha & \sin\alpha\\ -\sin\alpha & \cos\alpha\end{bmatrix},\quad \forall\alpha\in\mathscr{R}$$

为实正交矩阵. 若 $x\in\mathbb{C}^n$ 满足 $\|x\|=1$,则 $U=I-2xx^*$ 为 Hermite 矩阵与酉矩阵.

证明 直接验证便得 $A_\alpha A_\alpha^{\mathrm{T}}=I$, $\forall\alpha\in\mathbb{R}$;$U=U^*$;
$UU^*=(I-2xx^*)(I-2xx^*)=I-4xx^*+4xx^*xx^*=I-4xx^*+4xx^*=I$.

□

我们指出,形如 $U=I-2xx^*$ $(x^*x=1)$的矩阵通常叫作**初等酉矩阵**(或初等

Hermite 矩阵,或反射矩阵). 对任意 $y\in\mathbb{C}^n$,$(I-2xx^*)y$ 与 y 关于 x 与正交的超平面对称. 例 10 中两类特殊的矩阵在数值分析中十分有用.

$M_n(\mathbb{C})$中的 Hermite 矩阵与酉矩阵可以通过所谓 **Cayley 变换**将它们联系在一起,确切地说我们有如下结论.

定理 11(Cayley 变换定理) 设 $A\in M_n(\mathbb{C})$为 Hermite 矩阵. 则由 A 的如下 Cayley 变换定义的 $U\in M_n(\mathbb{C})$:

$$U=(A+\mathrm{i}I)^{-1}(A-\mathrm{i}I)=(A-\mathrm{i}I)(A+\mathrm{i}I)^{-1} \tag{24}$$

是酉矩阵,且 $1\notin\sigma(U)$. 反之,若 $U\in M_n(\mathbb{C})$为酉矩阵,且 $1\notin\sigma(U)$,则由下式定义的 $A\in M_n(\mathbb{C})$:

$$A=\frac{1}{\mathrm{i}}(U-I)^{-1}(U+I)=\frac{1}{\mathrm{i}}(U+I)(U-I)^{-1} \tag{25}$$

是 Hermite 矩阵,且 A 的 Cayley 变换为 U.

在第三章 3.1.2 小节命题 3 中将给出定理 11 的证明.

习题 1.2.1

1. 设 $A\in M_n(\mathbb{R})$为正规矩阵,它有实特征值 $r_1,\cdots,r_k$ 与复特征值 $c_j\pm \mathrm{i}d_j\ (j=1,\cdots,p)$,$2p+k=n$. 试证:存在正交矩阵 $U\in M_n(\mathbb{R})$使得

$$U^{\mathrm{T}}AU=\mathrm{diag}[r_1,\cdots,r_k]+\sum_{j=1}^{p}\begin{bmatrix}c_j & -d_j\\ d_j & c_j\end{bmatrix}.$$

2. 证明定理 8.

3. 设 $A\in M_n(\mathbb{C})$为 Hermite 矩阵,且 $A\neq O$. 试证:$\mathrm{rank}A\geqslant(\mathrm{tr}A)^2/\mathrm{tr}A^2$,并且,等式成立当且仅当 $A=\alpha UU^*$,这里,$\alpha\in\mathbb{R}$ 与 $U=[u_1\ \ \cdots\ \ u_r]\in M_{n,r}(\mathbb{C})$,诸 u_j 彼此正交且有长度 1.

4. 试证:若 $A\in M_n(\mathbb{C})$为正规矩阵,$\mathrm{Re}A=\frac{1}{2}(A+A^*)$($A$ 的实部)与 $\mathrm{Im}A=\frac{1}{2\mathrm{i}}(A-A^*)$($A$ 的虚部),则 $\mathrm{Re}A$ 与 $\mathrm{Im}A$ 可交换. 反之,若 $\mathrm{Re}A$ 与 $\mathrm{Im}A$ 可交换,则 A 为正规矩阵.

5. 设 $A\in M_n(\mathbb{C})$为正规矩阵且 $\lambda\in\sigma(A)$. 试证:$\mathrm{Re}\lambda\in\sigma(\mathrm{Re}A)$与 $\mathrm{Im}\lambda\in\sigma(\mathrm{I_m}A)$.

6. 设 $A\in M_r(\mathbb{C})$为正规矩阵. 试证:$\mathbb{C}^n=\mathrm{Im}A\oplus\mathrm{Ker}A$.

7. 设 $H\in M_n(\mathbb{C})$为 Hermite 矩阵,有秩 r,$H_1,\cdots,H_n$ 为 H 的前主子阵. 试证:若 $d_k=\det H_k\neq0$,$1\leqslant k\leqslant r$,则 H 的负(正)特征值个数等于序列 $1,d_1,\cdots,d_r$ 符号的变号数(连号数)(**Jacobi 法则**).

1.2.2 Hermite 正定与正半定矩阵

一个 Hermite 矩阵 $A\in M_n(F)$叫作**正定(正半定)的**,记为 $A>O(A\geqslant O)$,假如 $\langle Ax,x\rangle>0(\langle Ax,x\rangle\geqslant0)$,$\forall\,0\neq x\in F^n$. 下面关于 Hermite 正定矩阵的基本定理给出这样矩阵的一些特征.

定理 1(Hermite 正定矩阵的基本定理) 设 $A\in M_n(\mathbb{C})$为 Hermite 矩阵(或 $A\in M_n(\mathbb{R})$为对称矩阵). 则下面诸命题彼此等价:

(1) A 为正定的.

(2) A 的所有主子阵的特征值均为正数.

(3) A 的所有主子式均为正数.

(4) A 的所有前主子式均为正数,即 $\det A[(1,2,\cdots,k)]>0,k=1,\cdots,n$(**矩阵正定性的 Sylvester 准则**).

(5) 存在一个非奇异下三角形矩阵 $\widetilde{L}$ 使得 $A=\widetilde{L}\widetilde{L}^*$.

(6) 存在一个非奇异矩阵 C 使得 $A=CC^*$.

(7) 对于 $k=1,\cdots,n$,A 的所有 k 阶主子式之和为正数.

(8) A 的所有特征值为正数.

(9) 存在酉矩阵 U(或实正交矩阵 U)与正对角矩阵 D 使得 $A=UDU^*$.

证明　我们先证明前六个命题的等价性,然后证明(3)⇒(7)⇒(8)⇒(9)⇒(6).

(1)⇒(2):设 $\mathscr{M}\neq\varnothing$ 为 $\mathscr{N}=\{1,\cdots,n\}$ 的子集,$\lambda\in\sigma(A[\mathscr{M}])$,$\boldsymbol{x}[\mathscr{M}]\neq\boldsymbol{0}$ 满足 $A[\mathscr{M}]\boldsymbol{x}[\mathscr{M}]=\lambda\boldsymbol{x}[\mathscr{M}]$. 令 $\boldsymbol{x}\in\mathbb{C}^n$ 使得 $x_i=0,\forall i\notin\mathscr{M}$,而脚标属于 $\mathscr{M}$ 的分量组成的向量恰为 $\boldsymbol{x}[\mathscr{M}]$. 这时,$\langle A[\mathscr{M}]\boldsymbol{x}[\mathscr{M}],\boldsymbol{x}[\mathscr{M}]\rangle=\langle A\boldsymbol{x},\boldsymbol{x}\rangle>0$,因为 $\boldsymbol{x}\neq\boldsymbol{0}$. 于是,$\lambda\|\boldsymbol{x}[\mathscr{M}]\|^2>0$,从而 $\lambda>0$.

(2)⇒(3):这是显然的,因为 A 的任一主子式 $\det A[\mathscr{M}]$($\varnothing\neq\mathscr{M}\subset\mathscr{N}$)等于 $A[\mathscr{M}]$的所有特征值之积.

(3)⇒(4):显然成立.

(4)⇒(5):根据 1.1.1 小节定理 5,条件(4)蕴涵 A 有唯一的三角分解:$A=LU$,其中 L 为下三角形矩阵,对角元素均为 1,U 为非奇异的上三角形矩阵,对角元素均为正数. 这时,$A=A^*=U^*L^*=\widetilde{U}^*D^*L^*$,其中,$D$ 为正对角矩阵使得$U=D\widetilde{U}$($\widetilde{U}$ 为对角元素均等于 1 的上三角形矩阵). 按三角分解的唯一性可得,$L=\widetilde{U}^*$ 与 $U=D^*L^*=DL^*$. 令 W 为满足 $W^2=D$ 的正对角矩阵,且令 $\widetilde{L}=LW$. 则有$A=LU=(LW)(WL^*)=\widetilde{L}\widetilde{L}^*$,显然 $\widetilde{L}$ 为非奇异下三角形矩阵.

(5)⇒(6):显然成立.

(6)⇒(1):设 $A=CC^*$,C 为非奇异矩阵. 若 $\boldsymbol{0}\neq\boldsymbol{x}\in\mathbb{C}^n$,则 $C^*\boldsymbol{x}=\boldsymbol{y}\neq\boldsymbol{0}$,因而

$$\langle A\boldsymbol{x},\boldsymbol{x}\rangle=\langle CC^*\boldsymbol{x},\boldsymbol{x}\rangle=\|C^*\boldsymbol{x}\|^2=\|\boldsymbol{y}\|^2>0.$$

(3)⇒(7):显然成立.

(7)⇒(8):因 A 的所有特征值为实数,所以若有 A 的某特征值$\lambda\leqslant0$,按 1.1.1 小节(49)式则有

$$\det(\lambda I-A)=\sum_{t=0}^{n}\lambda^{n-t}(-1)^tE_t(A)=0,$$

其中,$E_0(A)=1$,$E_t(A)$为 A 的所有 t 阶主子式之和. 这是不可能的,因为 $\lambda\leqslant0$ 且 $E_t(A)>0,\forall t\geqslant1$.

(8)⇒(9)：按 1.2.1 小节定理 1(或定理 2)，有酉矩阵(或实正交矩阵)U 使得 $A=UDU^*$，这里 D 为对角矩阵，其对角元素恰为 A 的所有特征值. 这时条件(8)保证 D 为正对角矩阵.

(9)⇒(6)：设 $A=UDU^*$，这里 $D=\operatorname{diag}[d_1,\cdots,d_n]$，各 $d_j>0$. 现令 $W=\operatorname{diag}[\sqrt{d_1},\cdots,\sqrt{d_n}]$，则有 $D=W^2$ 与 $W=W^*$，因而 $A=(UW)(UW)^*$，显然，$C\equiv UW$ 非奇异. □

对于 Hermite 正半定矩阵，我们有如下结论.

定理 2(Hermite 正半定矩阵的基本定理)　设 $A\in\boldsymbol{M}_n(\mathbb{C})$ 为 Hermite 矩阵(或 $A\in\boldsymbol{M}_n(\mathbb{R})$ 为对称矩阵). 则下面诸命题彼此等价：

(1) A 为正半定的.

(2) $A+\varepsilon I>O$, $\forall\varepsilon>0$.

(3) A 的所有主子阵的特征值均为非负数.

(4) A 的所有主子式均为非负数.

(5) 对 $k=1,\cdots,n$，A 的所有 k 阶主子式之和为非负数.

(6) A 的所有特征值非负.

(7) 存在酉矩阵 U(或实正交矩阵 U)与非负对角矩阵 D 使得 $A=UDU^*$.

(8) 存在 $C\in\boldsymbol{M}_n(\mathbb{C})$(或 $\boldsymbol{M}_n(\mathbb{R})$)使得 $A=CC^*$.

本定理证明留给读者(见习题 1).

显然，任意 Hermite 正定矩阵也是正半定的. 下面定理告诉我们，非奇异的 Hermite 正半定矩阵必定为正定矩阵.

定理 3　设 $A\in\boldsymbol{M}_n(\mathbb{C})$ 为 Hermite 正半定矩阵. 则 A 为正定的当且仅当 A 非奇异.

证明　按定理 1(3)，A 为 Hermite 正定矩阵蕴涵 A 非奇异. 反之，若 A 为非奇异的 Hermite 正半定矩阵，则有 $\det A\neq0$，因而 A 的所有特征值非零，按定理 2(6)，这些特征值实际上均为正数. 再应用定理 1(8)便知 A 为正定的. □

定理 4　设 $A\in\boldsymbol{M}_n(\mathbb{C})$ 为 Hermite 矩阵. 则 A 为正定(正半定)的当且仅当存在唯一的 Hermite 正定(正半定)矩阵 W 使得 $W^2=A$. 此时，$\operatorname{rank}W=\operatorname{rank}A$.

证明　若 $A>O(\geqslant O)$，则有酉矩阵 U 使得 $A=UDU^*$，其中，$D=\operatorname{diag}[\lambda_1,\cdots,\lambda_n]$，$\lambda_j\in\sigma(A)$，$j=1,\cdots,n$ 且诸 $\lambda_j>0(\geqslant0)$. 现令 $D_0=\operatorname{diag}[\sqrt{\lambda_1},\cdots,\sqrt{\lambda_n}]$ 与 $W=UD_0U^*$. 则 $D_0>O(\geqslant O)$ 与 $W>O(\geqslant O)$ 且 $D_0^2=D$，$W^2=(UD_0U^*)(UD_0U^*)=UDU^*=A$. 现在证明 W 的唯一性. 设 $W>O(\geqslant O)$ 满足 $W^2=A$，且有谱形式 $W=\sum\limits_{k=1}^{p}\xi_kF_k$，则 $W^2=\sum\limits_{k=1}^{p}\xi_k^2F_k=A\equiv\sum\limits_{j=1}^{s}\lambda_jE_j$. 由于诸 $\xi_k\geqslant0$ 彼此不同，故诸 ξ_k^2 亦然，因而由 Hermite 矩阵 A 谱形式的唯一性推出，每个 ξ_k^2 必等于某个 λ_j，反之亦

然，于是，$p=s$，且 F_k 也等于对应的 E_j. 因此，$W=\sum_{j=1}^{s}\lambda_j^{1/2}E_j$. 此时有 $\text{rank}W=\text{rank}A$. □

Hermite 正定（正半定）矩阵还有许多重要的特征与有用的性质. 例如，当 $A\in\boldsymbol{M}_n(\mathbb{C})$ 为 Hermite 矩阵时，A 正定当且仅当 A^{-1} 正定. 若 A 为 Hermite 正定矩阵，则 A 的每一个主子阵也是 Hermite 正定的，且若 A_{11} 为 A 的主子阵，则在 A 内 A_{11} 的 Schur 余量 $[A/A_{11}]>O$. 事实上，不妨设 A 有对称分块矩阵形式：

$$A=\begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix},\qquad A_{11}\in\boldsymbol{M}_r(\mathbb{C}),\tag{1}$$

且 A^{-1} 也有类似的对称分块矩阵形式：

$$A^{-1}=\begin{bmatrix}B_{11} & B_{12}\\ B_{21} & B_{22}\end{bmatrix},\qquad B_{11}\in\boldsymbol{M}_r(\mathbb{C}).$$

根据 1.1.1 小节(29)式，我们有 $B_{22}^{-1}=[A/A_{11}]$. 再应用 Hermite 正定矩阵 A^{-1} 的主子阵 B_{22}（因而 B_{22}^{-1}）的正定性便得 $[A/A_{11}]>O$.

应用刚才证明的事实我们可以得到 Hermite 正定矩阵的又一特征.

定理 5　设 $A\in\boldsymbol{M}_n(\mathbb{C})$ 为 Hermite 矩阵，它有对称分块矩阵形式(1). 则 A 正定当且仅当 $A_{11}>O$ 与 $[A/A_{11}]>O$.

证明留给读者（见习题 2 或见后面定理 10 的证明）.

所有 n 阶 Hermite 正半定矩阵的集合 $\mathfrak{M}$ 还有一个值得注意的关于加法与数乘运算的封闭性质：若 $A,B\in\mathfrak{M}$ 与 $0\leqslant\alpha\in\mathbb{R}$，则 $A+B\in\mathfrak{M}$ 与 $\alpha A\in\mathfrak{M}$. 这表明所有 n 阶实对称正半定矩阵的集合 $\mathfrak{S}$ 组成 n 阶实对称矩阵的实线性空间内的一个凸锥. 从 1.2.1 小节定理 6 看出，$A\in\mathfrak{S}$ 当且仅当 A 的谱形式中所有 $\lambda_j\geqslant 0$. 这直接表明，$\mathfrak{S}$ 还是零矩阵 O_n 与所有秩 1 的 n 阶实对称正半定矩阵组成的集合的凸包.

现在讨论矩阵的**奇异值分解与极分解定理**.

Hermite 正半定矩阵 $H\in\boldsymbol{M}_n(\boldsymbol{F})$ 按定理 4 有唯一的 Hermite 正半定平方根（记为 $H^{1/2}$）的事实允许我们考察非方阵的谱特征. 设 $A\in\boldsymbol{M}_{m,n}(\boldsymbol{F})$，则 A^*A 与 AA^* 分别为 n 阶与 m 阶 Hermite 正半定矩阵，这是因为 $\langle A^*A\boldsymbol{x},\boldsymbol{x}\rangle=\|A\boldsymbol{x}\|^2\geqslant 0,\ \forall\,\boldsymbol{x}\in\boldsymbol{F}^n$；$\langle AA^*\boldsymbol{y},\boldsymbol{y}\rangle=\|A^*\boldsymbol{y}\|^2\geqslant 0,\ \forall\,\boldsymbol{y}\in\boldsymbol{F}^m$. 于是，$A^*A$ 与 AA^* 分别有 Hermite 正半定平方根 $H_1\equiv(A^*A)^{1/2}$ 与 $H_2\equiv(AA^*)^{1/2}$. 通常将 $H_1=(A^*A)^{1/2}\in\boldsymbol{M}_n(\boldsymbol{F})$ 的（非负）特征值 $\lambda_1,\lambda_2,\cdots,\lambda_n$ 称为 $A\in\boldsymbol{M}_{m,n}(\boldsymbol{F})$ 的**奇异值**（singular value），记为 $s_1,s_2,\cdots,s_n$，即 A 的奇异值 s_i 等于 $(A^*A)^{1/2}$ 的特征值 λ_i：

$$s_i(A)=\lambda_i((A^*A)^{1/2}),\ i=1,\cdots,n.\tag{2}$$

我们指出，有时也把 $H_2=(AA^*)^{1/2}\in\boldsymbol{M}_m(\boldsymbol{F})$ 的（非负）特征值 $\mu_1,\mu_2,\cdots,\mu_m$ 称为 $A\in\boldsymbol{M}_{m,n}(\boldsymbol{F})$ 的**奇异值**，这是因为按 1.1.1 小节(52)式，H_1 与 H_2 有相同的正特征值.

特别地，若 $A\in \boldsymbol{M}_n(\boldsymbol{F})$，则有 $s_i(A)=s_i(A^*)$，$i=1,\cdots,n$，并且，A 的奇异值在酉变换之下不变，即有

$$s_i(A)=s_i(UAV),i=1,\cdots,n, \tag{3}$$

式中，U 与 V 为任意的 n 阶酉矩阵. 这是因为按(2)式，$s_i(UAV)=\lambda_i((V^*A^*U^*UAV)^{1/2})=\lambda_i((V^*A^*AV)^{1/2})=\lambda_i((A^*A)^{1/2})=s_i(A)$. 当 $A\in \boldsymbol{M}_n(\boldsymbol{F})$ 为正规矩阵时，从 1.2.1 小节定理 1 容易看出，

$$s_i(A)=|\lambda_i(A)|,i=1,\cdots,n. \tag{4}$$

定理 6(奇异值分解定理)　设 $A\in \boldsymbol{M}_{m,n}(\boldsymbol{F})$ 有秩 $r>0$，且 $s_1,\cdots,s_r$ 为 A 的非零奇异值. 则存在 m 阶酉矩阵 U 与 n 阶酉矩阵 V 使得

$$A=USV^*, \tag{5}$$

式中，$S\in \boldsymbol{M}_{m,n}(\boldsymbol{F})$ 的 (i,i) 位置上元素等于 $s_i(1\leqslant i\leqslant r)$，其他元素均为零. 不计 $s_i(1\leqslant i\leqslant r)$ 的次序，S 是唯一确定的.

证明　不失一般性，假定 $m\leqslant n$，否则可考虑 A^*. 令 $\widetilde{A}=\begin{bmatrix}A\\O\end{bmatrix}\in \boldsymbol{M}_n(\boldsymbol{F})$，则有 $A^*A=\widetilde{A}^*\widetilde{A}\geqslant O$. 根据定理 2(7)，存在 n 阶酉矩阵 V 与非负对角矩阵 $D=\mathrm{diag}[d_1,\cdots,d_n]$ 使得 $A^*A=VDV^*$，因而

$$(AV)^*(AV)=D. \tag{6}$$

我们可以假定 $D=\mathrm{diag}[d_1,\cdots,d_p,0,\cdots,0]$，$d_j>0$，$j=1,\cdots,p$，并记 $\boldsymbol{w}_j=(AV)^{(j)}/\sqrt{d_j}$，$j=1,\cdots,p$. 按(6)式，诸 $\boldsymbol{w}_j$ 有长度 1 且彼此正交，因而存在 $\boldsymbol{w}_{p+1},\cdots,\boldsymbol{w}_m\in \boldsymbol{F}^m$ 使得 $U\equiv[\boldsymbol{w}_1\ \boldsymbol{w}_2\cdots\ \boldsymbol{w}_m]\in \boldsymbol{M}_m(\boldsymbol{F})$ 为酉矩阵. 现在证明 $AV=US$，其中，$S\in \boldsymbol{M}_{m,n}(\boldsymbol{F})$，它的 (i,i) 位置上元素等于 $\sqrt{d_i}$，$1\leqslant i\leqslant p$，其他元素均为零. 事实上，按 $\boldsymbol{w}_1,\cdots,\boldsymbol{w}_p$ 的定义，$AV=US$ 对 AV 的前 p 列显然成立，且按(6)式，AV 的其他 $n-p$ 列均为零向量. 因此，$A=USV^*$. 按 1.1.1 小节(37)式，$\mathrm{rank}A=\mathrm{rank}S=p=r$. $S\in \boldsymbol{M}_{m,n}(\boldsymbol{F})$ 的唯一性可从当 $1\leqslant i\leqslant p$ 时它的 (i,i) 位置上的元素恰为 $(A^*A)^{1/2}$ 的正的特征值的事实直接看到. □

作为上述定理的推论我们可以得出 n 阶矩阵的极分解. 它可视为复数的极形式：$\lambda=|\lambda|\mathrm{e}^{\mathrm{i}\theta}(0\leqslant\theta<2\pi)$ 的一种推广.

定理 7(矩阵的极分解)　设 $A\in \boldsymbol{M}_n(\boldsymbol{F})$. 则存在 $H\geqslant O$ 与 n 阶酉矩阵 U 使得

$$A=HU. \tag{7}$$

并且，(7)式中 H 唯一地等于 $(AA^*)^{1/2}$. 此外，假若 A 非奇异，那么 $H=(AA^*)^{1/2}>O$，这时，$U=H^{-1}A$，因而 A 的极分解 $A=HU$ 为唯一的. 类似地，存在 $H_1\geqslant O$ 与 n 阶酉矩阵 U_1 使得

$$A=U_1H_1, \tag{8}$$

式中，$H_1=(A^*A)^{1/2}$.

证明　显然,(8)式可从(7)式(对 A^*)直接推得. 按定理 6,存在 n 阶酉矩阵 U_1,V_1 与非负对角矩阵 S 使得 $A=U_1SV_1{}^*$. 现令 $H=U_1SU_1^*$ 与 $U=U_1V_1^*$,则 $H\geqslant O$,U 为 n 阶酉矩阵使得 $A=HU$,且 $AA^*=U_1S^2U_1^*=H^2$,因而 $H=(AA^*)^{1/2}$. 余下结论这时显然成立. □

在本小节最后部分,我们简略地介绍关于 Hermite 正定(正半定)矩阵的不等式.

在前面对 Hermite 矩阵 $A\in \boldsymbol{M}_n(\boldsymbol{F})$用记号 $A>O(\geqslant O)$表示 A 为正定(正半定)矩阵. 更一般地,若 $A,B\in \boldsymbol{M}_n(\boldsymbol{F})$为 Hermite 矩阵,则用记号 $A>B(A\geqslant B)$表示 $A-B>O(A-B)\geqslant O)$. 显然,这时"$\geqslant$"定义 n 阶 Hermite 矩阵之间的一种**偏序**(partial order). 下面考虑 Hermite 正半定矩阵的序性质.

首先容易看出,若 $A,B\in \boldsymbol{M}_n(\boldsymbol{F})$为 Hermite 矩阵,则 $A\geqslant B$ 蕴涵 $T^*AT\geqslant T^*BT$, $\forall T\in \boldsymbol{M}_{n,m}(\boldsymbol{F})$;$A>B$ 蕴涵 $T^*AT>T^*BT$, $\forall T\in \boldsymbol{M}_{n,m}(\boldsymbol{F})$但 $\operatorname{rank}T=m$.

其次,我们有如下结果.

定理 8　设 $A,B\in \boldsymbol{M}_n(\mathbb{C})$为 Hermite 矩阵,且 $A>O$ 与 $B\geqslant O$. 则 $A\geqslant B$ 当且仅当 $\rho(BA^{-1})\leqslant 1$;$A>B$ 当且仅当 $\rho(BA^{-1})<1$,这里 $\rho(C)$表示方阵 C 的**谱半径**,即 C 的特征值的最大模.

证明　按习题 4 的结果,有非奇异矩阵 $P\in \boldsymbol{M}_n(\mathbb{C})$使得 $A=PIP^*$ 与 $B=PDP^*$,这里 $D=\operatorname{diag}[d_1,\cdots,d_n]$为非负对角矩阵. 这时,$A\geqslant B$ 等价于 $P(I-D)\cdot P^*\geqslant O$,而后者又等价于 $d_i\leqslant 1, i=1,\cdots,n$. 但由于 $BA^{-1}=PDP^{-1}$,故 BA^{-1}与 D 有相同的特征值 $d_1,\cdots,d_n$,因而 $\rho(BA^{-1})=\max\limits_i d_i\leqslant 1$ 当且仅当所有 $d_i\leqslant 1$. 类似地可证明余下的结论. □

推论 9　若 $A,B\in \boldsymbol{M}_n(\mathbb{C})$为 Hermite 正定矩阵,则有:

(1) $A\geqslant B$ 当且仅当 $B^{-1}\geqslant A^{-1}$.

(2) $A\geqslant B$ 蕴涵 $\det A\geqslant \det B$, $\operatorname{tr}A\geqslant \operatorname{tr}B$,并且,若 A 与 B 的特征值分别按自大到小次序排列,则有 $\lambda_k(A)\geqslant\lambda_k(B)$, $k=1,\cdots,n$.

证明　(1) 按定理 8,$A\geqslant B$ 当且仅当 $\rho(BA^{-1})\leqslant 1$,而 $\rho(A^{-1}B)\leqslant 1$ 等价于 $B^{-1}\geqslant A^{-1}$. 但 $\rho(BA^{-1})=\rho(A^{-1}B)$.

(2) 若 $A\geqslant B$,则 $\rho(BA^{-1})\leqslant 1$. 按定理 8 的证明,BA^{-1}的特征值均非负,因而从 BA^{-1}非奇异便得 $\sigma(BA^{-1})\subset(0,1]$. 这表明 $\det(BA^{-1})=\det B\cdot\det A^{-1}\leqslant 1$ 进而有 $\det A\geqslant \det B$. 再按定理 8 的证明,有非奇异矩阵 $P=(p_{ij})\in \boldsymbol{M}_n(\mathbb{C})$使得 $A=PP^*$ 与 $B=PDP^*$, $D=\operatorname{diag}[d_1,\cdots,d_n]$, $0\leqslant d_i\leqslant 1, i=1,\cdots,n$. 由直接计算看出

$$\operatorname{tr}A=\operatorname{tr}(PP^*)=\sum_{i,j=1}^{n}|p_{ij}|^2$$

与

$$\mathrm{tr}B=\mathrm{tr}(PDP^*)=\mathrm{tr}(DP^*P)=\sum_{i,j=1}^{n}d_i\mid p_{ij}\mid^2\leqslant\sum_{i,j=1}^{n}\mid p_{ij}\mid^2=\mathrm{tr}A.$$

余下的结果直接从 1.2.1 小节定理 7 推出. □

定理 10　设 Hermite 矩阵 $A\in\boldsymbol{M}_n(\mathbb{C})$有对称分块形式

$$A=\begin{bmatrix}A_{11} & A_{12}\\ A_{12}^* & A_{22}\end{bmatrix},\qquad A_{11}\in\boldsymbol{M}_r(\mathbb{C}).$$

则 $A>O$ 当且仅当 $A_{11}>O$ 与 $A_{22}>A_{12}^*A_{11}^{-1}A_{12}$，并且，这后两个条件又等价于 $\rho(A_{12}^*A_{11}^{-1}A_{12}A_{22}^{-1})<1$，$A_{11}>O$ 与 $A_{22}>O$.

证明　若 $A>O$，我们有 $A_{11}>O$. 并且，$A^{-1}>O$. 按 1.1.1 小节(29)式(其中 $A_{21}=A_{12}^*$)，$(A_{22}-A_{12}^*A_{11}^{-1}A_{12})^{-1}>O$，此等价于 $A_{22}-A_{12}^*A_{11}^{-1}A_{12}>O$. 反之，设 $A_{11}>O$ 与 $A_{22}-A_{12}^*A_{11}^{-1}A_{12}>O$. 按 1.1.1 小节(28)式，我们容易看出 $A>O$. 最后的结果由定理 8 直接推出，因为在条件 $A_{22}>O$ 与 $A_{12}^*A_{11}^{-1}A_{12}\geqslant O$ 下，$A_{22}>A_{12}^*A_{11}^{-1}A_{12}$等价于 $\rho(A_{12}^*A_{11}^{-1}A_{12}A_{22}^{-1})<1$. □

现在介绍关于 Hermite 正定矩阵的 **Hadamard 不等式**.

定理 11(Hadamard 不等式)　设 $A=(a_{ij})\in\boldsymbol{M}_n(\mathbb{C})$为 Hermite 正定矩阵. 则有

$$\det A\leqslant\prod_{i=1}^{n}a_{ii},\tag{9}$$

等式成立当且仅当 A 为对角矩阵.

证明　令 $d_j=1/\sqrt{a_{jj}}(j=1,\cdots,n)$与 $D=\mathrm{diag}[d_1,\cdots,d_n]$. 由于 $\det(DAD)\leqslant1$ 等价于 $\det A\leqslant a_{11}a_{22}\cdots a_{nn}$，故我们可以假定所有 $a_{jj}=1$. 这时，若 $\lambda_1,\cdots,\lambda_n$ 为 A 的(正)特征值，则

$$\det A=\lambda_1\lambda_2\cdots\lambda_n\leqslant\left(\frac{1}{n}\sum_{j=1}^{n}\lambda_j\right)^n=\left(\frac{1}{n}\mathrm{tr}A\right)^n=1.$$

并且，上述不等式当且仅当 $\lambda_1=\cdots=\lambda_n=1$ 时变为等式. 但由于 A 为 Hermite 矩阵，按 1.2.1 小节定理 6，这种情形当且仅当 $A=I_n$ 时出现. 因此(9)式中等式成立当且仅当 A 为对角矩阵. □

对于一般方阵的另一个 Hadamard 不等式(见习题 1.1.1 题 9(2))，它本质上等价于 Hadamard 不等式(9).

推论 12(Hadamard 不等式)　设 $B=(b_{ij})\in\boldsymbol{M}_n(\mathbb{C})$. 则

$$\mid\det B\mid\leqslant\prod_{i=1}^{n}\left(\sum_{j=1}^{n}\mid b_{ij}\mid^2\right)^{1/2},\tag{10}$$

$$\mid\det B\mid\leqslant\prod_{j=1}^{n}\left(\sum_{i=1}^{n}\mid b_{ij}\mid^2\right)^{1/2},\tag{11}$$

(10)式中等式成立当且仅当 B 的 n 个行向量彼此正交；(11)式中等式成立当且仅

当 B 的 n 个列向量彼此正交.

证明 只要考虑不等式(10),因为用 B^* 代替 B 即得(11)式. 假如 B 奇异,(10)式显然成立. 假如 $\det B\neq 0$,对于 Hermite 正定矩阵 $A=BB^*$ 应用定理 11 直接便得不等式(10). B 的 n 个行向量彼此正交显然等价于 A 为对角矩阵,由定理 11 便知它为(10)式等式成立的充分必要条件. □

对 $A=(a_{ij}),B=(b_{ij})\in \boldsymbol{M}_{m,n}(\mathbb{C})$,矩阵 $C=(a_{ij}b_{ij})\in \boldsymbol{M}_{m,n}(\mathbb{C})$ 叫作 **A 与 B 的 Hadamard 积**,记为 $A\circ B$. 显然有 $A\circ B=B\circ A$. 对于两个同阶的 Hermite 正半定矩阵的 Hadamard 积有如下著名的 **Oppenheim 不等式**(12),它显然是定理 11 中 Hadamard 不等式(9)的一种推广.

定理 13(Oppenheim 不等式) 设 $A=(a_{ij}),B=(b_{ij})\in \boldsymbol{M}_n(\mathbb{C})$ 为 Hermite 正半定矩阵. 则 $A\circ B\geqslant O$,并且,

$$(\det A)\prod_{j=1}^{n} b_{jj}\leqslant \det(A\circ B). \tag{12}$$

证明 先证 $A,B\in \boldsymbol{M}_n(\mathbb{C})$ 的正半定性蕴涵 $A\circ B\geqslant O$. 根据 1.2.1 小节定理 6 及其证明,$A\geqslant O$ 与 $B\geqslant O$ 均可表示为秩 1 的 Hermite 正半定矩阵之和,于是按习题 14,有 $\boldsymbol{x}^{(p)}=(x_1^{(p)},\cdots,x_n^{(p)})^{\mathrm{T}},1\leqslant p\leqslant s$,与 $\boldsymbol{y}^{(q)}=(y_1^{(q)},\cdots,y_n^{(q)})^{\mathrm{T}},1\leqslant q\leqslant r$,使得

$$A=\sum_{p=1}^{s}\boldsymbol{x}^{(p)}(\boldsymbol{x}^{(p)})^* \quad 与 \quad B=\sum_{q=1}^{r}\boldsymbol{y}^{(q)}(\boldsymbol{y}^{(q)})^*,$$

从而有

$$a_{ij}=\sum_{p=1}^{s}x_i^{(p)}\bar{x}_j^{(p)} \quad 与 \quad b_{ij}=\sum_{q=1}^{r}y_i^{(q)}\bar{y}_j^{(q)},\quad 1\leqslant i,j\leqslant n.$$

因此,$A\circ B$ 的(i,j)元素可表示为

$$a_{ij}b_{ij}=\sum_{p=1}^{s}\sum_{q=1}^{r}x_i^{(p)}y_i^{(q)}\overline{(x_j^{(p)}y_j^{(q)})}.$$

由于同阶 Hermite 正半定矩阵之和仍是 Hermite 正半定矩阵,故为证 $A\circ B\geqslant O$,只须验证:对固定的 p 与 q,(i,j)位置上元素为 $x_i^{(p)}y_i^{(q)}\overline{(x_j^{(p)}y_j^{(q)})}$的 n 阶 Hermite 矩阵是正半定的. 令 $\boldsymbol{z}=(x_1^{(p)}y_1^{(q)},\cdots,x_n^{(p)}y_n^{(q)})^{\mathrm{T}}$,则这个 n 阶 Hermite 矩阵恰等于 $\boldsymbol{z}\boldsymbol{z}^*$,后者显然为正半定的. 因此,$A\circ B\geqslant O$.

其次证明不等式(12). 对阶数 n 应用数学归纳法. 当 $n=1$ 时,(12)式显然正确. 假定(12)式对阶数$\leqslant n-1$ 的所有 Hermite 正半定矩阵成立. 现考虑 n 阶情形. 令 $A_{11}\equiv A[(1,2,\cdots,n-1)]$与 $B_{11}\equiv B[(1,2,\cdots,n-1)]$,则因 $A_{11}\geqslant O$ 与 $B_{11}\geqslant O$,按归纳假定我们有

$$(\det A_{11})\prod_{j=1}^{n-1}b_{jj}\leqslant \det(A_{11}\circ B_{11}), \tag{13}$$

其中,$A_{11}\circ B_{11}=(A\circ B)[(1,2,\cdots,n-1)]$. 现令 $\tilde{A}=(\tilde{a}_{ij})\in \boldsymbol{M}_n(\mathbb{C})$使得 $\tilde{a}_{nn}=a_{nn}-$

α，其余 $\tilde{a}_{ij}=a_{ij}$，这里当 $A_{11}>O$ 时，$\alpha=\det A/\det A_{11}$，否则 $\alpha=0$. 这表明 A 与 $\tilde{A}$ 至多(n,n)元素有所不同. 这时，若 $A\geqslant O$ 为奇异的，则 $\tilde{A}=A$，否则有 $\det\tilde{A}=\det A-\alpha\det A_{11}=0$. 因此，不管 $A>O$ 或 $A\geqslant O$，$\tilde{A}$ 总是(奇异的)Hermite 正半定矩阵. 于是，$\tilde{A}\circ B\geqslant O$，因而 $\det(\tilde{A}\circ B)\geqslant 0$，亦即

$$\det(A\circ B)-\alpha b_{nn}\det(A_{11}\circ B_{11})\geqslant 0.$$

应用(13)式，我们有

$$\det(A\circ B)\geqslant \alpha b_{11}\cdots b_{nn}\det A_{11}=(\det A)\prod_{j=1}^{n}b_{jj}.$$ □

我们指出，从定理 13 直接推出，若 $A,B\in \boldsymbol{M}_n(\mathbb{C})$ 为 Hermite 正定矩阵，则 $A\circ B>O$. 这是因为按(12)式，$\det(A\circ B)\geqslant(\det A)\prod_{i=1}^{n}b_{ii}>0$(实际上，$A>O,B\geqslant O$ 但所有 $b_{ii}\neq 0$ 也蕴涵 $A\circ B>O$)

应用定理 11 与定理 13，我们可推出如下结果.

推论 14　设 $A=(a_{ij}),B=(b_{ij})\in \boldsymbol{M}_n(\mathbb{C})$ 为 Hermite 正定矩阵. 则

$$(\det A)(\det B)\leqslant \det(A\circ B), \tag{14}$$

$$(\det A)(\det B)\leqslant(\det A)\prod_{i=1}^{n}b_{ii}\leqslant\det(A\circ B)\leqslant\prod_{i=1}^{n}(a_{ii}b_{ii}), \tag{15}$$

$$1\leqslant\det(A\circ A^{-1}). \tag{16}$$

下面定理可以视为复数不等式 $|z|\geqslant \mathrm{Re}z(\forall z\in\mathbb{C})$ 的一种推广.

定理 15(Ostrowski-Taussky)　设 $A\in \boldsymbol{M}_n(\mathbb{C})$ 使得 $\mathrm{Re}A\equiv(A+A^*)/2>O$. 则

$$\det\mathrm{Re}A\leqslant|\det A|, \tag{17}$$

等式成立当且仅当 $A=A^*$.

证明　令 $\mathrm{Im}A\equiv(A-A^*)/(2\mathrm{i})$ 为 A 的虚部. 则 $A=\mathrm{Re}A+\mathrm{i}\mathrm{Im}A$，因而 $A=\mathrm{Re}A(I+(\mathrm{Re}A)^{-1}\mathrm{i}\mathrm{Im}A)$. 于是，我们只要证明：

$$|\det(I+(\mathrm{Re}A)^{-1}\mathrm{i}\mathrm{Im}A)|\geqslant 1.$$

但$(\mathrm{Re}A)^{-1}\mathrm{i}\mathrm{Im}A$ 相似$(\mathrm{Re}A)^{-1/2}\mathrm{i}\mathrm{Im}A(\mathrm{Re}A)^{-1/2}$，后者为斜 Hermite 矩阵，因而它们只有纯虚数的特征值，记为 $\mathrm{i}\alpha_1,\cdots,\mathrm{i}\alpha_n(\alpha_1,\cdots,\alpha_n\in\mathbb{R})$. 这样一来，

$$|\det(I+(\mathrm{Re}A)^{-1}\mathrm{i}\mathrm{Im}A)|=\prod_{j=1}^{n}|1+\mathrm{i}\alpha_j|\geqslant 1,$$

上式当且仅当所有 $\alpha_j=0$ 时等式成立. 因此，(17)式中等式成立当且仅当 $\mathrm{Im}A=O$ 即 $A=A^*$，因为 A 的虚部 $\mathrm{Im}A$ 为 Hermite 矩阵. □

另一个有关行列式的不等式涉及两个 Hermite 正定矩阵之和，它的证明类似于前一定理.

定理 16(Minkowski 不等式)　设 $A,B\in \boldsymbol{M}_n(\mathbb{C})$ 为 Hermite 正定矩阵. 则

$$(\det A)^{1/n}+(\det B)^{1/n}\leqslant(\det(A+B))^{1/n}. \tag{18}$$

证明 显然,不失普遍性可设 $A=I$,即只要证:

$$1+(\det B)^{1/n}\leqslant(\det(I+B))^{1/n}.$$

若 $0<\lambda_n\leqslant\cdots\leqslant\lambda_1$ 是 B 的特征值,则上述不等式等价于

$$(1+\sqrt[n]{\lambda_1\cdots\lambda_n})^n\leqslant\prod_{j=1}^{n}(1+\lambda_j).$$

但将上式两端乘开,并用算术-几何平均不等式逐项比较便能验明它是成立的. □

习题 1.2.2

1. 证明定理 2.

2. 证明定理 5.

3. 试证:若 $A\in\boldsymbol{M}_n(\mathbb{C})$ 为 Hermite 正定矩阵,P 为非奇异矩阵,则 $PAP^*>O$. 若 $A\geqslant O$ 且 $G\in\boldsymbol{M}_{m,n}(\mathbb{C})$,则 $GAG^*\geqslant O$.

4. 设 $A,B\in\boldsymbol{M}_n(\mathbb{C})$ 为 Hermite 矩阵,且对某 $\alpha,\beta\in\mathbb{R}$,$\alpha A+\beta B>O$. 试证:存在 n 阶非奇异矩阵 P 使得 P^*AP 与 P^*BP 均为对角矩阵. 特殊地,若 $A>O$,B 为 Hermite 矩阵($A,B\in\boldsymbol{M}_n(\mathbb{C})$),则存在非奇异矩阵 $P\in\boldsymbol{M}_n(\mathbb{C})$ 使得 $P^*AP=I$ 与 P^*BP 为对角矩阵.

5. 试证:当 A 取在 n 阶 Hermite 正定矩阵集合 $\mathfrak{M}$ 内时,$f(A)=\ln(\det A)$ 定义一个严格凹函数,即对所有 $\alpha\in(0,1)$,

$$f(\alpha A+(1-\alpha)B)\geqslant\alpha f(A)+(1-\alpha)f(B),\ \forall A,B\in\mathfrak{M},$$

并且,等式成立当且仅当 $A=B$.

6. 试证:若 $A,B\in\boldsymbol{M}_n(\mathbb{C})$ 为 Hermite 正定矩阵,且 $\alpha\in(0,1)$,则有

$$\det(\alpha A+(1-\alpha)B)\geqslant(\det A)^{\alpha}(\det B)^{1-\alpha},$$

等式成立当且仅当 $A=B$.

7. 试证:$f(A)=\mathrm{tr}A^{-1}$ 在 $\mathfrak{M}$ 上为严格凸函数,这里,$\mathfrak{M}$ 涵义同题 5.

8. 试证:若 $A,B\in\boldsymbol{M}_n(\mathbb{C})$ 为 Hermite 矩阵,$A>O$,则 $A+B>O$ 当且仅当 $A^{-1}B$ 的每一个特征值大于 -1.

9. 设 $A\in\boldsymbol{M}_n(\mathbb{C})$ 与 $C\in\boldsymbol{M}_m(\mathbb{C})$ 为 Hermite 正定矩阵,$B\in\boldsymbol{M}_{n,m}(\mathbb{C})$. 试证下列命题彼此等价:

(1) $(\boldsymbol{x}^*A\boldsymbol{x})(\boldsymbol{y}^*C\boldsymbol{y})\geqslant|\boldsymbol{x}^*B\boldsymbol{y}|^2,\ \forall\boldsymbol{x}\in\mathbb{C}^n,\boldsymbol{y}\in\mathbb{C}^m$.

(2) $\boldsymbol{x}^*A\boldsymbol{x}+\boldsymbol{y}^*C\boldsymbol{y}\geqslant2|\boldsymbol{x}^*B\boldsymbol{y}|,\ \forall\boldsymbol{x}\in\mathbb{C}^n,\boldsymbol{y}\in\mathbb{C}^m$.

(3) $\rho(B^*A^{-1}BC^{-1})\leqslant1$.

(4) $\begin{bmatrix}A & B\\ B^* & C\end{bmatrix}\geqslant O$.

10. 在前题中当 $A=C$ 时,证明下列命题等价:

(1) $(\boldsymbol{x}^*A\boldsymbol{x})(\boldsymbol{y}^*A\boldsymbol{y})\geqslant|\boldsymbol{x}^*B\boldsymbol{y}|^2,\ \forall\boldsymbol{x},\boldsymbol{y}\in\mathbb{C}^n$.

(2) $\boldsymbol{x}^*A\boldsymbol{x}+\boldsymbol{y}^*A\boldsymbol{y}\geqslant2|\boldsymbol{x}^*B\boldsymbol{y}|^2,\ \forall\boldsymbol{x},\boldsymbol{y}\in\mathbb{C}^n$.

(3) $\rho(B^*A^{-1}BA^{-1})\leqslant1$.

(4) $\boldsymbol{x}^*A\boldsymbol{x}\geqslant|\boldsymbol{x}^*B\boldsymbol{x}|,\ \forall\boldsymbol{x}\in\mathbb{C}^n$.

11. 设 $P\in\boldsymbol{M}_n(\mathbb{C})$ 为 Hermite 正定矩阵,它有对称分块矩阵形式:

$$P=\begin{bmatrix} A & B \\ B^* & C \end{bmatrix},\quad A\in \boldsymbol{M}_r(\mathbb{C}),\quad 0<r<n.$$

试证如下所谓 **Fischer 不等式**成立：

$$\det P \leqslant (\det A)(\det C).$$

12. 设 $A,B\in \boldsymbol{M}_n(\mathbb{C})$ 为 Hermite 正定矩阵. 试证：

$$(\det A)\prod_{j=1}^{n} b_{jj} + (\det B)\prod_{j=1}^{n} a_{jj} \leqslant \det(A\circ B)+(\det A)(\det B).$$

13. 设 $A\in \boldsymbol{M}_n(\mathbb{C})(n\geqslant 2)$ 使得 $\mathrm{Re}A>O$. 试证定理 15 的如下加强结果：

$$\det \mathrm{Re}A + |\det \mathrm{Im}A| \leqslant |\det A|.\ (\mathrm{Im}A \text{ 为 } A \text{ 的虚部})$$

14. 设 $A=(a_{ij})\in \boldsymbol{M}_n(\mathbb{C})$ 为 Hermite 矩阵，且 $\mathrm{rank}A=1$. 试证：存在 $\boldsymbol{x}\in\mathbb{C}^n$ 与 $\alpha\in\mathbb{R}$ 使得 $A=\alpha \boldsymbol{x}\boldsymbol{x}^*$. 特殊地，若 $A\geqslant O$，$\mathrm{rank}A=1$，则存在 $\boldsymbol{y}\in\mathbb{C}^n$ 使得 $A=\boldsymbol{y}\boldsymbol{y}^*$.

15. 设 $A_0,A_1,\cdots,A_m\in \boldsymbol{M}_n(\mathbb{C})$ 为 Hermite 矩阵. 试证：若 $A_0>O$，则存在 $\varepsilon>0$ 使得 $B(x)=A_0+xA_1+\cdots+x^mA_m>O$，$\forall x\in[-\varepsilon,\varepsilon]$.

1.2.3　幂等变换与幂等矩阵

回顾一下，$P\in \boldsymbol{M}_n(\boldsymbol{F})$ 叫作幂等矩阵，假如 $P^2=P$. 按此定义我们有，$P^k=P$，$k=1,2,\cdots$，且各幂等矩阵的谱包含在集合 $\{0,1\}$ 内. 类似地，对带有内积 $\langle,\rangle$ 的酉空间 $\mathscr{S}$ 上的线性变换 T，若 $T^2=T$，则称 $T\in\mathscr{L}(\mathscr{S})$ 为**幂等变换**. 幂等变换通常也叫作**投影**(理由见后面). 显然，若 $T\in\mathscr{L}(\mathscr{S})$ 为幂等变换，则 $\mathscr{A}_T^0$(见 1.2.1 小节)中任意矩阵都是幂等矩阵. 本小节主要讨论幂等矩阵与 Hermite 幂等矩阵的有关结果. 对于幂等变换也能找到类似的结论.

幂等矩阵有许多重要的性质，首先有下列结果.

定理 1　设 $P\in \boldsymbol{M}_n(\boldsymbol{F})$ 为幂等矩阵，则

(1) P^* 与 $I-P$ 为幂等矩阵.

(2) $\sigma(P)\subset\{0,1\}$.

(3) $\mathrm{Im}(I-P)=\mathrm{Ker}P$，$\mathrm{Ker}(I-P)=\mathrm{Im}P$.

(4) $\mathrm{Ker}P\dot{+}\mathrm{Im}P=\boldsymbol{F}^n$.

(5) P 为简单矩阵，$\mathrm{rank}P=\mathrm{tr}P$.

证明　(1)与(2)显然成立. 若 $\boldsymbol{y}\in\mathrm{Im}(I-P)$，则对某 $\boldsymbol{x}\in\boldsymbol{F}^n$，$\boldsymbol{y}=(I-P)\boldsymbol{x}$，因而 $P\boldsymbol{y}=(P-P^2)\boldsymbol{x}=\boldsymbol{0}$，即 $\boldsymbol{y}\in\mathrm{Ker}P$. 反之，若 $P\boldsymbol{y}=\boldsymbol{0}$，则 $(I-P)\boldsymbol{y}=\boldsymbol{y}$，因而 $\boldsymbol{y}\in\mathrm{Im}(I-P)$. 于是，$\mathrm{Im}(I-P)=\mathrm{Ker}P$. 用 $I-P$ 换 P 可证 $\mathrm{Ker}(I-P)=\mathrm{Im}P$. 因此(3)也成立. 为证(4)，设 $\boldsymbol{x}\in\boldsymbol{F}^n$，则 $\boldsymbol{x}=\boldsymbol{x}_1+\boldsymbol{x}_2$，这里 $\boldsymbol{x}_1=(I-P)\boldsymbol{x}\in\mathrm{Ker}P$ 与 $\boldsymbol{x}_2=P\boldsymbol{x}\in\mathrm{Im}P$，这表明 $\boldsymbol{F}^n=\mathrm{Ker}P+\mathrm{Im}P$. 若 $\boldsymbol{x}\in\mathrm{Ker}P\cap\mathrm{Im}P$，则 $P\boldsymbol{x}=\boldsymbol{0}$ 与 $\boldsymbol{x}\in\mathrm{Im}P=\mathrm{Ker}(I-P)$，因而 $(I-P)\boldsymbol{x}=\boldsymbol{0}$. 由此得出 $\boldsymbol{x}=P\boldsymbol{x}=\boldsymbol{0}$，于是，$\mathrm{Ker}P+\mathrm{Im}P$ 为直和. 由于任意 $\boldsymbol{0}\neq\boldsymbol{x}\in\mathrm{Im}P$ 满足 $\boldsymbol{x}=P\boldsymbol{x}$，即 $\boldsymbol{x}$ 为对应于特征值 1 的 P 的特征向量，同时，任意 $\boldsymbol{0}\neq\boldsymbol{y}\in\mathrm{Ker}P$ 满足 $P\boldsymbol{y}=\boldsymbol{0}$，即 $\boldsymbol{y}$ 为对应于特征值 0 的 P 的特征向量，故由(2)与

(4)知道,我们有 P 的 n 个线性无关的特征向量,因而 P 为简单矩阵.显然,这时 $\mathrm{rank}P$ 等于特征值 1 的代数重数,即有 $\mathrm{rank}P=\mathrm{tr}P$,(5)得证. □

我们指出,定理 1(4)表明每一个幂等矩阵 P 生成 $\boldsymbol{F}^n$ 中互补子空间 $\mathscr{S}_1=\mathrm{Ker}P$ 与 $\mathscr{S}_2=\mathrm{Im}P$.反之,若 $\mathscr{S}_1\dotplus\mathscr{S}_2=\boldsymbol{F}^n$,则存在唯一的幂等矩阵 P 使得 $\mathrm{Ker}P=\mathscr{S}_1$ 与 $\mathrm{Im}P=\mathscr{S}_2$.因为这时任意 $\boldsymbol{x}\in\boldsymbol{F}^n$ 可唯一地表示为

$$\boldsymbol{x}=\boldsymbol{x}_1+\boldsymbol{x}_2,\quad \boldsymbol{x}_1\in\mathscr{S}_1 \text{ 与 } \boldsymbol{x}_2\in\mathscr{S}_2, \tag{1}$$

定义 $P:\boldsymbol{F}^n\to\boldsymbol{F}^n$ 使得 $P\boldsymbol{x}=\boldsymbol{x}_2$,则显然有 $P^2=P$,$\mathrm{Ker}P=\mathscr{S}_1$ 与 $\mathrm{Im}P=\mathscr{S}_2$,且 P 的唯一性直接由此定义得出.因此,$\boldsymbol{M}_n(\boldsymbol{F})$内的幂等矩阵与 $\boldsymbol{F}^n$ 内互补子空间对 $\mathscr{S}_1,\mathscr{S}_2$ 之间有着一一对应关系.由于这个原因,可以将幂等矩阵 P 看成为**沿着 $\mathrm{Ker}P$ 在 $\mathrm{Im}P$ 上的投影**.同样地,由定理 1 看出,幂等矩阵 $I-P$ 便是沿着 $\mathrm{Im}P$ 在 $\mathrm{Ker}P$ 上的投影.此外,当且仅当 $\boldsymbol{x}\in\mathrm{Im}P$ 时,$P\boldsymbol{x}=\boldsymbol{x}$;当且仅当 $\boldsymbol{x}\in\mathrm{Ker}P$ 时,$P\boldsymbol{x}=\boldsymbol{0}$.

基于定理 1(2)与(5),应用简单矩阵的谱定理(1.1.1 小节定理 7)我们有关于幂等矩阵的特殊结果.

定理 2(幂等矩阵的谱定理)　设 $P\in\boldsymbol{M}_n(\boldsymbol{F})$为幂等矩阵.则有 $\boldsymbol{F}^n$ 内的双正交系$\{\boldsymbol{x}_1,\cdots,\boldsymbol{x}_r\}$,$\{\boldsymbol{y}_1,\cdots,\boldsymbol{y}_r\}$(即有$\langle\boldsymbol{x}_i,\boldsymbol{y}_j\rangle=\boldsymbol{y}_j^*\boldsymbol{x}_i=\delta_{ij}$,$1\leqslant i,j\leqslant r$)使得

$$P=\sum_{j=1}^{r}\boldsymbol{x}_j\boldsymbol{y}_j^*, \tag{2}$$

这里,$r=\mathrm{rank}P$.

我们指出,直接计算表明具有形式(2)的矩阵 P 满足 $P=P^2$,因而是幂等矩阵.注意,(2)式中各 $\boldsymbol{x}_j\boldsymbol{y}_j^*$ 为秩 1 矩阵.

现在考察同阶幂等矩阵的代数组合仍为幂等矩阵的条件.下列定理的证明留给读者(见习题 1).

定理 3　设 $P_1,P_2\in\boldsymbol{M}_n(\boldsymbol{F})$为幂等矩阵.则

(1) 当且仅当 $P_1P_2=P_2P_1=O$ 时,$P_1+P_2\in\boldsymbol{M}_n(\boldsymbol{F})$为幂等矩阵.此时,$P\equiv P_1+P_2$ 为沿着 $\mathrm{Ker}P_1\cap\mathrm{Ker}P_2$ 在 $\mathrm{Im}P_1\dotplus\mathrm{Im}P_2$ 上的投影.

(2) 当且仅当 $P_1P_2=P_2P_1=P_2$ 时,$P_1-P_2\in\boldsymbol{M}_n(\boldsymbol{F})$为幂等矩阵.此时,$Q\equiv P_1-P_2$ 为沿着 $\mathrm{Ker}P_1\dotplus\mathrm{Im}P_2$ 在 $\mathrm{Im}P_1\cap\mathrm{Ker}P_2$ 上的投影.

(3) 若 $P_1P_2=P_2P_1=R$,则 R 为沿着 $\mathrm{Ker}P_1+\mathrm{Ker}P_2$ 在 $\mathrm{Im}P_1\cap\mathrm{Im}P_2$ 上的投影.

我们指出,与(1),(2)情形不同,定理 3(3)中出现子空间之和 $\mathrm{Ker}P_1+\mathrm{Ker}P_2$ 而不是直和 $\mathrm{Ker}P_1\dotplus\mathrm{Ker}P_2$.只要考虑到 $P_1=P_2=P$ 的可能性,这种现象就不难理解.

从前面讨论看出,幂等矩阵或幂等变换与 $\boldsymbol{F}^n$ 的直和分解有着密切的联系,因而我们可应用幂等矩阵或幂等变换来研究 1.1.2 小节中线性变换及其矩阵表示的可约性与完全可约性问题.

定理 4　设 $\mathscr{S}_1$ 为 $\mathscr{S}$ 的线性子空间. 若 $\mathscr{S}_1$ 为 $T\in\mathscr{L}(\mathscr{S})$ 的不变子空间,则对 $\mathscr{S}_1$ 上所有幂等变换 P 有 $PTP=TP$. 反之,若对 $\mathscr{S}_1$ 上的某个幂等变换 P 有 $PTP=TP$,则 $\mathscr{S}_1$ 为 T-不变的.

证明　任取某 $\mathscr{S}_2$ 使得 $\mathscr{S}=\mathscr{S}_1\dot{+}\mathscr{S}_2$,且设 P 为沿着 $\mathscr{S}_2$ 在 $\mathscr{S}_1$ 上的投影变换. 这时,对任意 $\boldsymbol{z}=\boldsymbol{x}+\boldsymbol{y}\in\mathscr{S}(\boldsymbol{x}\in\mathscr{S}_1$ 与 $\boldsymbol{y}\in\mathscr{S}_2)$,我们有 $TP\boldsymbol{z}=T\boldsymbol{x}$,因而 $PTP\boldsymbol{z}=PT\boldsymbol{x}=T\boldsymbol{x}$,因为 $\boldsymbol{x}\in\mathscr{S}_1$ 蕴涵 $T\boldsymbol{x}\in\mathscr{S}_1$. 这表明 $PTP\boldsymbol{z}=TP\boldsymbol{z}$, $\forall\boldsymbol{z}\in\mathscr{S}$,于是,$PTP=TP$. 反之,假定对沿着某 $\mathscr{S}_2$ 在 $\mathscr{S}_1$ 上的投影变换 P 有 $PTP=TP$,且 $\boldsymbol{x}\in\mathscr{S}_1$,则有 $P\boldsymbol{x}=\boldsymbol{x}$,因而 $PT\boldsymbol{x}=PTP\boldsymbol{x}=TP\boldsymbol{x}=T\boldsymbol{x}$. 于是,$T\boldsymbol{x}\in\mathscr{S}_1$, $\forall\boldsymbol{x}\in\mathscr{S}_1$,即 $\mathscr{S}_1$ 为 T-不变的. □

定理 5　设 $\mathscr{S}_1$ 与 $\mathscr{S}_2$ 为 $\mathscr{S}$ 的两个线性子空间使得 $\mathscr{S}=\mathscr{S}_1\dot{+}\mathscr{S}_2$. 则 $\mathscr{S}_1$ 与 $\mathscr{S}_2$ 皆为 $T\in\mathscr{L}(\mathscr{S})$ 的不变子空间当且仅当 $PT=TP$,这里,P 为沿着 $\mathscr{S}_2$ 在 $\mathscr{S}_1$ 上的投影变换.

证明　若 $PT=TP$(P 为沿着 $\mathscr{S}_2$ 在 $\mathscr{S}_1$ 上的投影变换),且 $\boldsymbol{x}\in\mathscr{S}_1$ 与 $\boldsymbol{y}\in\mathscr{S}_2$,则 $T\boldsymbol{x}=TP\boldsymbol{x}=PT\boldsymbol{x}$ 与 $P\boldsymbol{y}=\boldsymbol{0}$,因而 $PT\boldsymbol{y}=TP\boldsymbol{y}=\boldsymbol{0}$. 于是,$T\boldsymbol{x}\in\mathscr{S}_1$, $\forall\boldsymbol{x}\in\mathscr{S}_1$, $T\boldsymbol{y}\in\mathscr{S}_2$, $\forall\boldsymbol{y}\in\mathscr{S}_2$. 反之,若 $\mathscr{S}_1$ 与 $\mathscr{S}_2$ 皆为 T-不变的,则定理 4 保证 $PTP=TP$ 与 $(I-P)T(I-P)=T(I-P)$,但后一式子可改写为 $PTP=PT$,因而 $PT=TP$. □

这个定理给出线性变换 $T\in\mathscr{L}(\mathscr{S})$ 可以表示为直和形式的等价条件. 当定理 5 条件满足时,适当选取 $\mathscr{S}$ 的基 $\mathscr{E}$,T 的矩阵表示 $[T]_{\mathscr{E}}$ 将呈分块对角矩阵形式. 同理,适当选取 $\mathscr{S}$ 的基 $\mathscr{E}$,在定理 4 条件下,T 的矩阵表示 $[T]_{\mathscr{E}}$ 将呈分块上三角形矩阵的形式.

现在考虑幂等变换 $P\in\mathscr{L}(\mathscr{S})$($\mathscr{S}$ 为酉空间)的伴随变换 P^*.

定理 6　若 P 为沿着 $\mathscr{S}_2$ 在 $\mathscr{S}_1$ 上的投影变换,$\mathscr{S}=\mathscr{S}_1\dot{+}\mathscr{S}_2$,则 P^* 便是沿着 $\mathscr{S}_1^{\perp}$ 在 $\mathscr{S}_2^{\perp}$ 上的投影变换.

证明　首先从定理 1 看出,P^* 为幂等变换. 余下只要证:$\mathrm{Im}P^*=\mathscr{S}_2^{\perp}$ 与 $\mathrm{Ker}P^*=\mathscr{S}_1^{\perp}$. 若 $\boldsymbol{y}\in\mathscr{S}_2^{\perp}$,则对任意 $\boldsymbol{x}\in\mathscr{S}$,$\langle\boldsymbol{x},\boldsymbol{y}\rangle=\langle P\boldsymbol{x},\boldsymbol{y}\rangle+\langle(I-P)\boldsymbol{x},\boldsymbol{y}\rangle=\langle P\boldsymbol{x},\boldsymbol{y}\rangle=\langle\boldsymbol{x},P^*\boldsymbol{y}\rangle$,因而 $\boldsymbol{y}=P^*\boldsymbol{y}$,这表明 $\mathscr{S}_2^{\perp}\subset\mathrm{Im}P^*$. 反之,若 $\boldsymbol{y}\in\mathrm{Im}P^*$,则 $\boldsymbol{y}=P^*\boldsymbol{y}$,因而对任意 $\boldsymbol{x}\in\mathscr{S}_2$,$\langle\boldsymbol{x},\boldsymbol{y}\rangle=\langle\boldsymbol{x},P^*\boldsymbol{y}\rangle=\langle P\boldsymbol{x},\boldsymbol{y}\rangle=0$,这表明 $\boldsymbol{y}\in\mathscr{S}_2^{\perp}$. 因此,$\mathrm{Im}P^*=\mathscr{S}_2^{\perp}$. 余下证明:$\mathrm{Ker}P^*=\mathscr{S}_1^{\perp}$. 若 $\boldsymbol{y}\in\mathrm{Ker}P^*$,则 $P^*\boldsymbol{y}=\boldsymbol{0}$,这时,$\langle\boldsymbol{x},\boldsymbol{y}\rangle=\langle P\boldsymbol{x},\boldsymbol{y}\rangle=\langle\boldsymbol{x},P^*\boldsymbol{y}\rangle=0$, $\forall\boldsymbol{x}\in\mathscr{S}_1$,因而 $\boldsymbol{y}\in\mathscr{S}_1^{\perp}$. 反之,若 $\boldsymbol{y}\in\mathscr{S}_1^{\perp}$,则对任意 $\boldsymbol{x}\in\mathscr{S}$,$\langle\boldsymbol{x},P^*\boldsymbol{y}\rangle=\langle P\boldsymbol{x},\boldsymbol{y}\rangle=0$,因而 $P^*\boldsymbol{y}=\boldsymbol{0}$ 即有 $\boldsymbol{y}\in\mathrm{Ker}P^*$. 因此,$\mathrm{Ker}P^*=\mathscr{S}_1^{\perp}$. □

应用定理 1 与定理 6,我们容易得到 Hermite 幂等矩阵(或自伴幂等变换)的以下等价命题.

定理 7　设 $P\in\boldsymbol{M}_n(\boldsymbol{F})$ 为幂等矩阵,则下列命题彼此等价:

(1) P 为 Hermite 矩阵,即 $P=P^*$.

(2) P 为正规矩阵,即 $P^*P=PP^*$.

(3) $\mathrm{Ker}P \oplus \mathrm{Im}P = \boldsymbol{F}^n$，即有 $\mathrm{Im}P = (\mathrm{Ker}P)^{\perp}$.

(4) $\langle P\boldsymbol{x}, \boldsymbol{x}\rangle = \| P\boldsymbol{x} \|^2, \forall \boldsymbol{x} \in \boldsymbol{F}^n$.

证明　(1)⇒(2)：显然成立.

(2)⇒(3)：对任意 $\boldsymbol{x} \in \mathrm{Ker}P$ 与 $\boldsymbol{y} \in \mathrm{Im}P$，$P$ 为正规矩阵蕴涵 $\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \langle \boldsymbol{x}, P^* P\boldsymbol{y}_1\rangle = \langle P\boldsymbol{x}, P\boldsymbol{y}_1\rangle = 0$（这里，$\boldsymbol{y}_1 \in \boldsymbol{F}^n$ 为某一合适向量），因为 $\mathrm{Im}P = \mathrm{Im}(PP^*) = \mathrm{Im}(P^*P)$，所以 $\boldsymbol{y} \in \mathrm{Im}P$ 等价于有 $\boldsymbol{y}_1 \in \boldsymbol{F}^n$ 使得 $\boldsymbol{y} = P^*P\boldsymbol{y}_1$. 因此，按定理 1(4)，$\mathrm{Im}P = (\mathrm{Ker}P)^{\perp}$.

(3)⇒(4)：对任意 $\boldsymbol{x} = \boldsymbol{y} + \boldsymbol{z} \in \boldsymbol{F}^n$，$\boldsymbol{y} \in \mathrm{Im}P$ 与 $\boldsymbol{z} \in \mathrm{Ker}P$，我们有 $\langle P\boldsymbol{x}, \boldsymbol{x}\rangle = \langle \boldsymbol{y}, \boldsymbol{y}+\boldsymbol{z}\rangle = \langle \boldsymbol{y}, \boldsymbol{y}\rangle = \| P\boldsymbol{x} \|^2$.

(4)⇒(1)：由(4)推出，$\langle (P-P^*)\boldsymbol{x}, \boldsymbol{x}\rangle = 0, \forall \boldsymbol{x} \in \boldsymbol{F}^n$，因而如令 $T = P - P^*$，则我们有 $\langle T(\boldsymbol{x}+\boldsymbol{y}), \boldsymbol{x}+\boldsymbol{y}\rangle = \langle T\boldsymbol{x}, \boldsymbol{x}\rangle + \langle T\boldsymbol{x}, \boldsymbol{y}\rangle + \langle T\boldsymbol{y}, \boldsymbol{x}\rangle + \langle T\boldsymbol{y}, \boldsymbol{y}\rangle = 0$，即有 $\langle T\boldsymbol{x}, \boldsymbol{y}\rangle + \langle T\boldsymbol{y}, \boldsymbol{x}\rangle = 0, \forall \boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{F}^n$. 用 $\mathrm{i}\boldsymbol{y}$ 代替前一式子中的 $\boldsymbol{y}$ 便得 $-\mathrm{i}\langle T\boldsymbol{x}, \boldsymbol{y}\rangle + \mathrm{i}\langle T\boldsymbol{y}, \boldsymbol{x}\rangle = 0, \forall \boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{F}^n$，于是 $\langle T\boldsymbol{x}, \boldsymbol{y}\rangle - \langle T\boldsymbol{y}, \boldsymbol{x}\rangle = 0, \forall \boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{F}^n$. 考虑到 $\langle T\boldsymbol{x}, \boldsymbol{y}\rangle + \langle T\boldsymbol{y}, \boldsymbol{x}\rangle = 0, \forall \boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{F}^n$ 的已有结论，因此我们有 $\langle T\boldsymbol{x}, \boldsymbol{y}\rangle = 0, \forall \boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{F}^n$. 取 $\boldsymbol{y} = T\boldsymbol{x}$ 便得 $\| T\boldsymbol{x} \|^2 = 0$，因而 $T\boldsymbol{x} = \boldsymbol{0}, \forall \boldsymbol{x} \in \boldsymbol{F}^n$，即有 $P = P^*$. □

我们指出，由于定理 7(3)，Hermite 幂等矩阵 P 又称为**正交投影矩阵**，它将 $\boldsymbol{F}^n$ 的元素沿着 $\mathrm{Ker}P$ 投射到 $\mathrm{Ker}P$ 的正交补空间 $\mathrm{Im}P$ 上. 由于 $(\mathrm{Im}P)^{\perp} = \mathrm{Ker}P$，这时我们可简称 P 为**在 $\mathrm{Im}P$ 上的正交投影**，记为 $P = P_{\mathrm{Im}P}$. 此外，从定理 7(4)看出，正交投影 P 一定为 Hermite 正半定矩阵，且有性质

$$\| P\boldsymbol{x} \| \leqslant \| \boldsymbol{x} \|, \forall \boldsymbol{x} \in \boldsymbol{F}^n. \tag{3}$$

事实上，由于 $I - P$ 也是正交投影，故有 $\| \boldsymbol{x} \|^2 - \| P\boldsymbol{x} \|^2 = \langle \boldsymbol{x}, \boldsymbol{x}\rangle - \langle P\boldsymbol{x}, \boldsymbol{x}\rangle = \langle (I-P)\boldsymbol{x}, \boldsymbol{x}\rangle = \| (I-P)\boldsymbol{x} \|^2 \geqslant 0, \forall \boldsymbol{x} \in \boldsymbol{F}^n$. 有趣的是，满足条件(3)的幂等矩阵 P 必定是 Hermite 矩阵（见习题 2）.

一般地，若酉空间 $\mathscr{S}$ 有正交和表示：$\mathscr{S} = \mathscr{T} \oplus \mathscr{T}^{\perp}$，则可以定义线性变换 $T \in \mathscr{L}(\mathscr{S})$ 使得 $T\boldsymbol{x} = \boldsymbol{y}$，这里 $\boldsymbol{x} = \boldsymbol{y} + \boldsymbol{z} \in \mathscr{S}, \boldsymbol{y} \in \mathscr{T}, \boldsymbol{z} \in \mathscr{T}^{\perp}$. 这样的变换 T 叫作**在 $\mathscr{T}$ 上的正交投影变换**，记为 $P_{\mathscr{T}}$. 这时，$T = T^*$，$\mathrm{Ker}T = \mathscr{T}^{\perp}$ 与 $\mathrm{Im}T = \mathscr{T}$，并且类似于定理 7 的结论也成立. 两个正交投影变换 T_1 与 T_2 叫作**正交的**，假如 $T_1T_2 = O$. 考虑它们的伴随变换便知此条件等价于 $T_2T_1 = O$.

定理 8　设 $T_1, T_2 \in \mathscr{L}(\mathscr{S})$ 为正交投影变换，且 $T_1 = P_{\mathscr{T}_1}$ 与 $T_2 = P_{\mathscr{T}_2}$. 则当且仅当 $\mathscr{T}_1$ 与 $\mathscr{T}_2$ 正交时，T_1 与 T_2 正交.

证明　若 $T_1T_2 = O$，且 $\boldsymbol{x} \in \mathscr{T}_1 = \mathrm{Im}T_1$ 与 $\boldsymbol{y} \in \mathscr{T}_2 = \mathrm{Im}T_2$ 则有 $\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \langle T_1\boldsymbol{x}, T_2\boldsymbol{y}\rangle = \langle \boldsymbol{x}, T_1^*T_2\boldsymbol{y}\rangle = \langle \boldsymbol{x}, T_1T_2\boldsymbol{y}\rangle = 0$，因而 $\mathscr{T}_1$ 与 $\mathscr{T}_2$ 正交. 反之，若 $\mathscr{T}_1$ 与 $\mathscr{T}_2$ 正交，则 $\mathscr{T}_2 \subset \mathscr{T}_1^{\perp}$，因而 $T_1\boldsymbol{x} = \boldsymbol{0}, \forall \boldsymbol{x} \in \mathscr{T}_1^{\perp}$，它蕴涵 $T_1T_2\boldsymbol{x} = \boldsymbol{0}, \forall \boldsymbol{x} \in \mathscr{S}$，因为 $T_2\boldsymbol{x} \in \mathscr{T}_2$，所以 $T_2\boldsymbol{x} \in \mathscr{T}_1^{\perp}$. □

回顾一下 1.2.1 小节中的正规矩阵与 Hermite 矩阵的谱定理（见 1.2.1 小节

定理 5 与 6)，其中，诸 G_j 皆为秩 1 的 Hermite 幂等矩阵，且两两正交；诸 E_j 也是两两正交的 Hermite 幂等(非零)矩阵.

由于正交投影矩阵即是 Hermite 幂等矩阵，故按幂等矩阵的谱定理(定理 2)与 Hermite 矩阵的谱定理(1.2.1 小节定理 6)我们有如下结论.

定理 9(正交投影矩阵的谱定理)　设 $P\in \boldsymbol{M}_n(\boldsymbol{F})$ 为正交投影矩阵，$\mathrm{rank}P=r$，$\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n$ 为 P 的标准正交特征向量，其中，$\boldsymbol{x}_1,\cdots,\boldsymbol{x}_r$ 对应于特征值 1，其余的对应于零特征值. 则有

$$P=\sum_{j=1}^{r}\boldsymbol{x}_j\boldsymbol{x}_j^*. \tag{4}$$

并且，

$$P\boldsymbol{x}=\sum_{j=1}^{r}\langle \boldsymbol{x},\boldsymbol{x}_j\rangle \boldsymbol{x}_j,\ \forall\, \boldsymbol{x}\in \boldsymbol{F}^n. \tag{5}$$

习题 1.2.3

1. 证明定理 3.

2. 试证：满足条件(3)的幂等矩阵 P 为 Hermite 矩阵.

3. 试证：若 $T_1,\cdots,T_m$ 为 $\mathscr{L}(\mathscr{S})$ 中的正交投影变换，则 $T=T_1+\cdots+T_m\in\mathscr{L}(\mathscr{S})$ 为正交投影变换当且仅当诸 T_j 两两正交，即 $T_iT_j=O,\ \forall\, i\neq j$.

4. 设 $T_1=P_{\mathscr{T}_1}$ 与 $T_2=P_{\mathscr{T}_2}$ 分别为酉空间 $\mathscr{S}$ 的子空间 $\mathscr{T}_1$ 与 $\mathscr{T}_2$ 上的正交投影变换. 试证下列诸命题彼此等价：

(1) T_1-T_2 为正交投影变换.

(2) $T_1\geqslant T_2$，即 $\langle T_1\boldsymbol{x},\boldsymbol{x}\rangle\geqslant\langle T_2\boldsymbol{x},\boldsymbol{x}\rangle,\ \forall\,\boldsymbol{x}\in\mathscr{S}$.

(3) $\|T_1\boldsymbol{x}\|\geqslant\|T_2\boldsymbol{x}\|,\ \forall\,\boldsymbol{x}\in\mathscr{S}$.

(4) $\mathscr{T}_2\subset\mathscr{T}_1$.

(5) $T_1T_2=T_2$.

(6) $T_2T_1=T_2$.

5. 设 $P\in\boldsymbol{M}_n(\mathbb{C})$. 试证：$P$ 为正交投影矩阵当且仅当 $P=P^*P$.

参 考 文 献

[1] Marcus M, Minc H. A Survey of Matrix Theory and Matrix Inequalities. New York: Dover Publica- tions, 1992

[2] Lancaster P, Tismenetsky M. The Theory of Matrices with Applications. 2nd ed. New York: Academic Press, 1985

[3] Halmos P R. Finite-Dimensional Vector Spaces. Princeton: Van Nostrand, N. J. 1958

[4] Fielder M. Special Matrices and Their Applications in Numerical Mathematics. Dordrecht Martinus Nijhoff Publishers, 1986

[5] Horn R A, Johnson C R. Matrix Analysis . Cambridge: Cambridge Univ. Press, 1994

[6] Bellman R. Introduction to Matrix Analysis. New York: McGraw-Hill, 1960; Philadelphia, SIAM Press, 1995

[7] Hoffman K, Kunze R. Linear Algebra. 2nd ed. Prentice-Hall, Englewood Cliffs, N. J. , 1971

[8] Mirsky L. An Introduction to Linear Algebra. Oxford: Clarendon Press, 1963

[9] 甘特马赫尔 Φ P. 矩阵论(上下册). 柯召译. 北京:高等教育出版社,1955

[10] 盖尔冯德 U M. 线性代数学. 刘亦珩译. 北京:高等教育出版社,1957

[11] 北京大学数力系几何与代数教研室代数小组. 高等代数. 北京:人民教育出版社,1978

[12] 张禾瑞,郝钢新. 高等代数. 北京:人民教育出版社,1979

[13] Bhatia R. Matrix Analysis. GTM169, New York: Springer-Velag, 1997

[14] Meyer C D. Matrix Aralysis and Applied Linear Algebra. Philadelphia: SIAM Press, 2000

第二章 范 数

2.1 向 量 范 数

在许多场合，对同一线性空间中的向量需要引入非负数量作为它们"大小"的一种度量，进而比较两个向量之间的"接近"程度. 我们希望这样的度量类似于复数的模或$\mathbb{R}^3$ 中的向量欧氏长度.

对于一般酉空间 $\mathscr{U}$，自然地可以应用内积$\langle \boldsymbol{x},\boldsymbol{y}\rangle$导出向量 $\boldsymbol{x}$ 的"长度"或范数(norm)：$\|\boldsymbol{x}\|=\langle \boldsymbol{x},\boldsymbol{x}\rangle^{1/2}$. 若 $\|\boldsymbol{x}\|$ 为小的数，则认为 $\boldsymbol{x}$(按此"长度")是"小"的向量. 同时，若 $\boldsymbol{x},\boldsymbol{y}\in\mathscr{U}$ 使得 $\|\boldsymbol{x}-\boldsymbol{y}\|$ 为小的数，则认为按此"长度"$\boldsymbol{x}$ 与 $\boldsymbol{y}$ 是彼此"接近"的向量. 但由本节后面讨论看出，范数不一定从内积导出，它只要满足一些基本的公理条件即可.

本节讨论一般有限维线性空间上的向量范数概念与范数的有关性质.

2.1.1 定义与例子

设 $\mathscr{S}$表示数域 $\boldsymbol{F}(=\mathbb{R}$ 或$\mathbb{C})$上的线性空间.

定义 1 $\mathscr{S}$上的**向量范数**(vector norm)是一个实值函数 $\|\cdot\|:\mathscr{S}\rightarrow\mathbb{R}$，它满足下列四个条件：

(1) $\|\boldsymbol{x}\|\geqslant 0,\forall\boldsymbol{x}\in\mathscr{S}$(非负性)；

(2) $\|\alpha\boldsymbol{x}\|=|\alpha|\ \|\boldsymbol{x}\|,\forall\boldsymbol{x}\in\mathscr{S}$与$\forall\alpha\in\boldsymbol{F}$(齐次性)；

(3) $\|\boldsymbol{x}+\boldsymbol{y}\|\leqslant\|\boldsymbol{x}\|+\|\boldsymbol{y}\|,\forall\boldsymbol{x},\boldsymbol{y}\in\mathscr{S}$(三角不等式)；

(4) 当且仅当 $\boldsymbol{x}=\boldsymbol{0}$ 时，$\|\boldsymbol{x}\|=0$(正性).

前面讲过的由酉空间内积$\langle\cdot,\cdot\rangle$引出的向量"长度"满足定义 1 中所有条件，因为按内积定义，条件(1)，(2)与(4)显然成立，为验证条件(3)，应用 **Cauchy-Schwarz 不等式**

$$|\langle\boldsymbol{x},\boldsymbol{y}\rangle|\leqslant\|\boldsymbol{x}\|\ \|\boldsymbol{y}\|,\tag{1}$$

于是，

$$\begin{aligned}\|\boldsymbol{x}+\boldsymbol{y}\|^2&=\langle\boldsymbol{x}+\boldsymbol{y},\boldsymbol{x}+\boldsymbol{y}\rangle\leqslant\|\boldsymbol{x}\|^2+2\|\boldsymbol{x}\|\ \|\boldsymbol{y}\|+\|\boldsymbol{y}\|^2\\&=(\|\boldsymbol{x}\|+\|\boldsymbol{y}\|)^2.\end{aligned}$$

通常，将这样的向量范数称为**由内积$\langle\cdot,\cdot\rangle$导出的向量范数**. 可以证明：$\mathscr{S}$上向量范数 $\|\cdot\|$ 是由内积$\langle\cdot,\cdot\rangle$导出的充分与必要条件为它满足**平行四边形恒等式**(见习题 4)：

$$\frac{1}{2}(\|x+y\|^2+\|x-y\|^2)=\|x\|^2+\|y\|^2. \tag{2}$$

假如 $\mathscr{S}$ 上实值函数 $\|\cdot\|:\mathscr{S}\to\mathbb{R}$ 满足定义 1 中条件(2)与(3)，则称它是 $\mathscr{S}$ 上的**向量拟范数或向量半模**(vector seminorm). 显然，它是向量范数概念的一种推广，因为它允许非零向量有零长度. 我们将在本小节末尾给出向量拟范数的例子.

定义 1 中条件(2)与(3)蕴涵 $\|\mathbf{0}\|=0$，条件(1)以及不等式：

$$|\,\|x\|-\|y\|\,|\leqslant\|x-y\|,\ \forall x,y\in\mathscr{S}. \tag{3}$$

因此，我们有(见习题 3)

引理 2 若 $\|\cdot\|$ 为 $\mathscr{S}$ 上的向量拟范数，则 $\|\mathbf{0}\|=0$，定义 1 中条件(1)与(3)式成立.

当 $\mathscr{S}=\mathbb{C}^n$(或 $\mathbb{R}^n$)时，下列实值函数都是 $\mathbb{C}^n$(或 $\mathbb{R}^n$)上的向量范数：

(1) $\|x\|_\infty\equiv\max\limits_{1\leqslant j\leqslant n}|x_j|$(**最大范数或 l_∞ 范数**)；

(2) $\|x\|_1\equiv\sum\limits_{j=1}^{n}|x_j|$(**和范数或 l_1 范数**) (4)

(3) $\|x\|_2\equiv(\sum\limits_{j=1}^{n}|x_j|^2)^{1/2}=(x^*x)^{1/2}$(**欧氏范数或 l_2 范数**)；

(4) $\|x\|_p\equiv(\sum\limits_{j=1}^{n}|x_j|^p)^{1/p},p\geqslant1$(**Hölder 范数或 l_p 范数**).

容易看出，$\|\cdot\|_1$ 与 $\|\cdot\|_2$ 为 Hölder 范数中取 $p=1$ 与 $p=2$ 的特殊情形，而 $\|\cdot\|_\infty$ 是 $\|\cdot\|_p$ 当 $p\to\infty$ 时的极限情形. 事实上，对 $x\neq\mathbf{0}$，令 $\|x\|_\infty=\rho$，且设 x 有 v 个分量 $x_{j_1},\cdots,x_{j_v}$ 的模同为 ρ. 则当 $p\to\infty$ 时，

$$\|x\|_p=(\sum_{j=1}^{n}|x_j|^p)^{1/p}=\rho(v+\sum_{i=v+1}^{n}(|x_{j_i}|/\rho)^p)^{1/p}\to\rho.$$

直接验证便知，$\|\cdot\|_\infty$ 满足定义 1 的四个定义，因而它确是 $\mathbb{C}^n$(或 $\mathbb{R}^n$)上的向量范数. 对于 Hölder 范数 $\|\cdot\|_p(p\geqslant1)$，它们满足定义 1 中条件(1)，(2)与(4)也是显然的，为了验证条件(3)，我们只要应用下列 **Minkowski 不等式**即可：

$$(\sum_{j=1}^{n}|x_j+y_j|^p)^{1/p}\leqslant(\sum_{j=1}^{n}|x_j|^p)^{1/p}+(\sum_{j=1}^{n}|y_j|^p)^{1/p},\quad p\geqslant1.$$

若 $\mathscr{S}$ 为一般 n 维线性空间，且 $\mathscr{B}=\{b_1,\cdots,b_n\}$ 为 $\mathscr{S}$ 的基，则对 $x=\sum\limits_{j=1}^{n}\alpha_jb_j\in\mathscr{S}$，应用

$$x\to(\alpha_1,\cdots,\alpha_n)^{\mathrm{T}}\in\mathbb{C}^n$$

是 $\mathscr{S}$ 到 $\mathbb{C}^n$ 上同构的事实，可以证明：如 $\|\cdot\|$ 为 $\mathbb{C}^n$ 上的向量范数，那么如下的实值函数 $\|\cdot\|_{\mathscr{B}}$：

$$\|x\|_{\mathscr{B}}\equiv\|(\alpha_1,\cdots,\alpha_n)^{\mathrm{T}}\| \tag{5}$$

为 $\mathscr{S}$ 上的向量范数. 因此，特殊地如下定义的 $\|\cdot\|_\infty$ 与 $\|\cdot\|\,p(p\geqslant1)$ 皆为 $\mathscr{S}$ 上的向量

范数：

$$\|x\|_\infty \equiv \max_{1\leqslant j\leqslant n} |\alpha_j|,$$

$$\|x\|_p \equiv \left(\sum_{j=1}^{n} |\alpha_j|^p\right)^{1/p}, \qquad p \geqslant 1. \tag{6}$$

并且，$\|x\|_\infty = \lim_{p\to\infty} \|x\|_p$，$\forall x \in \mathscr{S}$.

给定 $\mathscr{S}$ 上的一个或几个向量范数，我们可以按一些方法构造出 $\mathscr{S}$ 上新的向量范数. 例如，若 $\|\cdot\|_\alpha$ 与 $\|\cdot\|_\beta$ 为 $\mathscr{S}$ 上的向量范数，则它们之和 $\|\cdot\|_\alpha + \|\cdot\|_\beta$ 还是 $\mathscr{S}$ 上的向量范数，若 $r>0$，则 $r\|\cdot\|_\alpha$ 也是 $\mathscr{S}$ 上的向量范数；此外，由下式定义的实值函数 $\|\cdot\|$：

$$\|x\| \equiv \max\{\|x\|_\alpha, \|x\|_\beta\} \tag{7}$$

也是 $\mathscr{S}$ 上的向量范数. 更一般地我们有

定理 3 若 $\|\cdot\|_{\alpha_1}, \cdots, \|\cdot\|_{\alpha_m}$ 为 $\mathscr{S}$ 上的 m 个向量范数，且 $\|\cdot\|_\beta$ 为 $\mathbb{R}^m$ 上的单调向量范数（即若 $x, y \in \mathbb{R}^m$，且 $x-y$ 的所有分量非负，则有 $\|x\|_\beta \geqslant \|y\|_\beta$），那么实值函数 $\|\cdot\|: \mathscr{S} \to \mathbb{R}$：

$$\|x\| = \|(\|x\|_{\alpha_1}, \cdots, \|x\|_{\alpha_m})^{\mathrm{T}}\|_\beta \tag{8}$$

为 $\mathscr{S}$ 上的向量范数.

上述定理的证明留给读者（见习题 5）. 我们指出，若"向量范数"换为"向量拟范数"，则定理 3 仍成立（见习题 5）.

下面定理给出构造 $\mathbb{C}^n$ 上新向量范数的另一种方法.

定理 4 若 $\|\cdot\|$ 为 $\mathbb{C}^n$ 上向量范数，且 $A \in \boldsymbol{M}_n(\mathbb{C})$ 非奇异，则由下式定义的实值函数 $\|\cdot\|_A$：

$$\|x\|_A \equiv \|Ax\|, \qquad \forall x \in \mathbb{C}^n \tag{9}$$

也是 $\mathbb{C}^n$ 上的向量范数.

这个定理的证明也留给读者（见习题 6）. 应用定理 4，例如我们可以直接验证

$$\|x\| \equiv (|2x_1 - x_2|^2 + |x_2|^2)^{1/2}, \forall x \in \mathbb{C}^2$$

为 $\mathbb{C}^2$ 上的向量范数，因为这时有

$$\|x\| = \|Ax\|_2, \qquad A = \begin{bmatrix} 2 & -1 \\ 0 & 1 \end{bmatrix}.$$

当定理 4 中 $A \in \boldsymbol{M}_n(\mathbb{C})$ 奇异时，(9) 式定义的实值函数 $\|\cdot\|_A$ 只是 $\mathbb{C}^n$ 上的向量拟范数，此时有 $\mathbf{0} \neq x \in \mathbb{C}^n$ 使得 $\|x\|_A = 0$，因为这时 A 的零空间 $\mathrm{Ker}A$ 有非零的向量.

习题 2.1.1

1. 若 $H \in \boldsymbol{M}_n(\mathbb{C})$ 为 Hermite 正定矩阵，试证：

$$\|x\| \equiv (x^* Hx)^{1/2}$$

为$\mathbb{C}^n$上的向量范数(**椭圆范数**).

2. 试证:$\mathbb{C}^n$上的欧氏范数$\|\cdot\|_2$为**酉不变的**,即若$U\in \boldsymbol{M}_n(\mathbb{C})$为酉矩阵,则$\|U\boldsymbol{x}\|_2=\|\boldsymbol{x}\|_2$, $\forall \boldsymbol{x}\in\mathbb{C}^n$.

3. 证明引理 2.

4. 试证:$\mathscr{S}$上向量范数$\|\cdot\|$是由内积$\langle\cdot,\cdot\rangle$导出的充分与必要条件为它满足平行四边形恒等式(2).

5. 证明定理 3,并证明:若向量范数换为向量拟范数,则定理 3 仍成立.

6. 证明定理 4.

7. 试证:任一向量拟范数有形式(9),这里,$\|\cdot\|$为某向量范数,$A\in \boldsymbol{M}_n(\mathbb{C})$.

8. 设$\|\cdot\|$为$\mathbb{C}^n$上酉不变向量范数,试证:存在正数α使得$\|\cdot\|=\alpha\|\cdot\|_2$,且$\|\cdot\|_2$为满足条件$\|\boldsymbol{e}_1\|=1$的唯一酉不变向量范数,这里,$\boldsymbol{e}_1=(1,0,\cdots,0)^{\mathrm{T}}\in\mathbb{C}^n$.

9. 试证:

$$\|\boldsymbol{y}\|_\infty=\max_{\|\boldsymbol{x}\|_1=1}|\boldsymbol{y}^*\boldsymbol{x}|,\quad \|\boldsymbol{x}\|_1=\max_{\|\boldsymbol{y}\|_\infty=1}|\boldsymbol{x}^*\boldsymbol{y}|.$$

10. 试证:当$0<p<1$时,$\|\boldsymbol{x}\|_p=\left(\sum_{j=1}^n|x_j|^p\right)^{1/p}$不是$\mathbb{C}^n$(或$\mathbb{R}^n$)上的向量范数.

2.1.2 分析与几何性质

由前面讨论知道,有限维空间$\mathscr{S}$上可以定义许多不同的范数$\|\cdot\|$,它们组成的不同空间$(\mathscr{S},\|\cdot\|)$都称为**赋范线性空间**.赋范空间$(\mathscr{S},\|\cdot\|)$的拓扑可以用$\|\boldsymbol{x}-\boldsymbol{y}\|$定义的两个向量$\boldsymbol{x}$与$\boldsymbol{y}$之间的**距离**来确定.显然,它确实满足距离的公理条件:

(1) $\|\boldsymbol{x}-\boldsymbol{y}\|\geqslant 0$;当且仅当$\boldsymbol{x}=\boldsymbol{y}$时,$\|\boldsymbol{x}-\boldsymbol{y}\|=0$;

(2) $\|\boldsymbol{x}-\boldsymbol{y}\|=\|\boldsymbol{y}-\boldsymbol{x}\|$;

(3) $\|\boldsymbol{x}-\boldsymbol{y}\|\leqslant\|\boldsymbol{x}-\boldsymbol{z}\|+\|\boldsymbol{z}-\boldsymbol{y}\|$.

空间$(\mathscr{S},\|\cdot\|)$的**(开)球**定义为集合

$$B_{\|\cdot\|}(r;\boldsymbol{x}_0)=\{\boldsymbol{x}\in\mathscr{S}:\|\boldsymbol{x}-\boldsymbol{x}_0\|<r\},\tag{1}$$

式中,$\boldsymbol{x}_0$为球心,$r>0$为该球的半径.因此,$(\mathscr{S},\|\cdot\|)$的**单位(开)球**为$B_{\|\cdot\|}\equiv B_{\|\cdot\|}(1;\boldsymbol{0})$.类似地可以定义**闭球**与**单位闭球**分别为$\{\boldsymbol{x}\in\mathscr{S}:\|\boldsymbol{x}-\boldsymbol{x}_0\|\leqslant r\}$与$\{\boldsymbol{x}\in\mathscr{S}:\|\boldsymbol{x}\|\leqslant 1\}$.给定$\boldsymbol{x}\in\mathscr{S}$,$\mathscr{S}$的包含以$\boldsymbol{x}$为球心某球的子集都叫作$\boldsymbol{x}$的**邻域**.$\mathscr{S}$的子集叫作**有界的**,假如它包含在某闭球内.注意,除非$\dim\mathscr{S}=0$,$\mathscr{S}$本身不是有界的.

对于$\mathscr{S}$的任一子集S,$\boldsymbol{x}$称为S的**内点**,如果S为$\boldsymbol{x}$的邻域;$\boldsymbol{y}$称为S的**外点**,如果$\boldsymbol{y}$是S的余集$\mathscr{S}-S$的内点;z称为S的**边界点**,如果z既不是S的内点也不是S的外点.今后用∂S表示S的所有边界点集合.S与∂S的并集叫作S的**闭包**,记为$\overline{S}$.S为$\mathscr{S}$内**开集**,如果它仅由内点组成;S为$\mathscr{S}$内**闭集**,如果S的余集$\mathscr{S}-S$为开集,或等价地,如果$S=\overline{S}$.$\mathscr{S}$的任一子集S的闭包为闭集:$\overline{\overline{S}}=\overline{S}$.$\mathscr{S}$的任一子空间为闭的(因$\mathscr{S}$为有限维空间).

定义 1 赋范线性空间$(\mathscr{S}, \|\cdot\|)$的向量序列$\{\boldsymbol{x}^{(k)}\}$称为**收敛于** $\boldsymbol{x}\in\mathscr{S}$或**有极限** $\boldsymbol{x}$,假如

$$\lim_{k\to\infty}\|\boldsymbol{x}^{(k)}-\boldsymbol{x}\|=0. \tag{2}$$

任一有限维赋范空间内的收敛向量序列$\{\boldsymbol{x}^{(k)}\}$必定为 **Cauchy 序列**,即有$\|\boldsymbol{x}^{(k)}-\boldsymbol{x}^{(l)}\|\to 0, k,l\to\infty$,反之亦然(见习题 1). 因此,这样的空间为完备的. 这表明任一有限维赋范空间$(\mathscr{S}, \|\cdot\|)$为 Banach 空间.

当赋范线性空间$(\mathscr{S}, \|\cdot\|)$的向量序列有极限时,极限一定唯一. 同时,$\mathscr{S}$内的线性运算按如下意义为连续的:

若$\{\boldsymbol{x}^{(k)}\}$与$\{\boldsymbol{y}^{(k)}\}$分别有极限$\boldsymbol{x}$与$\boldsymbol{y}$,$\boldsymbol{F}$中数列$\{\mu^{(k)}\}$与$\{\nu^{(k)}\}$分别有极限μ与ν,则$\mu^{(k)}\boldsymbol{x}^{(k)}+\nu^{(k)}\boldsymbol{y}^{(k)}$有极限$\mu\boldsymbol{x}+\nu\boldsymbol{y}$.

前述的与赋范空间相联系的拓扑概念也可以用收敛序列来定义. 例如,S是闭集当且仅当$\boldsymbol{x}^{(k)}\in S$与$\boldsymbol{x}^{(k)}$有极限$\boldsymbol{x}$蕴涵$\boldsymbol{x}\in S$;$S$的闭包$\overline{S}=\{\boldsymbol{x}\in\mathscr{S}$:存在$\boldsymbol{x}^{(k)}\in S$使得$\boldsymbol{x}^{(k)}$有极限$\boldsymbol{x}\}$,等等.

对于**有限维**空间情形,下面的范数等价定理保证向量序列的收敛定义与所选用的$\mathscr{S}$上向量范数无关. 我们说,$\mathscr{S}$上两个向量范数$\|\cdot\|_\alpha$与$\|\cdot\|_\beta$是**等价的**(equivalent),如果存在只与$\|\cdot\|_\alpha$与$\|\cdot\|_\beta$有关的正常数c与d,使得

$$c\|\boldsymbol{x}\|_\alpha\leqslant\|\boldsymbol{x}\|_\beta\leqslant d\|\boldsymbol{x}\|_\alpha, \forall\boldsymbol{x}\in\mathscr{S}. \tag{3}$$

先证明一个引理,它表明$\mathscr{S}$上任一向量范数依坐标是连续的,即$\lim\limits_{k\to\infty}\|\boldsymbol{x}^{(k)}-\boldsymbol{x}\|_\infty=0$蕴涵$\|\boldsymbol{x}^{(k)}\|\to\|\boldsymbol{x}\|\ (k\to\infty)$.

引理 2 设$\|\cdot\|$为$\mathscr{S}$上任一向量范数. 则对任给$\varepsilon>0$,有$\delta>0$,只要$\|\boldsymbol{x}-\boldsymbol{y}\|_\infty<\delta$,便有

$$|\ \|\boldsymbol{x}\|-\|\boldsymbol{y}\|\ |<\varepsilon. \tag{4}$$

证明 设$\mathscr{B}=\{\boldsymbol{b}_1,\cdots,\boldsymbol{b}_n\}$为$\mathscr{S}$的基,令$M=\max\limits_{1\leqslant j\leqslant n}\|\boldsymbol{b}_j\|$,并取$\delta=\varepsilon M^{-1}n^{-1}$. 则对$\boldsymbol{x}=\sum\limits_{j=1}^n\alpha_j\boldsymbol{b}_j$与$\boldsymbol{y}=\sum\limits_{j=1}^n\beta_j\boldsymbol{b}_j$,只要$\|\boldsymbol{x}-\boldsymbol{y}\|_\infty<\delta$,便有

$$\begin{aligned}\|\boldsymbol{x}-\boldsymbol{y}\|&\leqslant\sum_{j=1}^n|\alpha_j-\beta_j|\ \|\boldsymbol{b}_j\|\\&\leqslant M\sum_{j=1}^n|\alpha_j-\beta_j|<Mn\delta=\varepsilon.\end{aligned}$$

应用 2.1.1 小节(3)式即得(4)式. □

定理 3(范数等价定理) 有限维线性空间$\mathscr{S}$上任意两个向量范数是等价的.

证明 我们只需证明:$\mathscr{S}$上任一向量范数$\|\cdot\|$均与$\mathscr{S}$上向量范数$\|\cdot\|_2$等价. 设$\mathscr{B}=\{\boldsymbol{b}_1,\cdots,\boldsymbol{b}_n\}$为$\mathscr{S}$的基. 按引理 2,考虑定义在$\boldsymbol{F}^n$上的实值连续函数$g$:

$$g(\alpha_1,\cdots,\alpha_n)=\left\|\sum_{j=1}^n\alpha_j\boldsymbol{b}_j\right\|,\qquad\forall(\alpha_1,\cdots,\alpha_n)^{\mathrm{T}}\in\boldsymbol{F}^n.$$

由于集合 $S=\{(\alpha_1,\cdots,\alpha_n)^{\mathrm{T}}\in \boldsymbol{F}^n:\ \sum_{j=1}^{n}|\alpha_j|^2=1\}$在 $\boldsymbol{F}^n$ 内为闭与有界的，故按分析中的 Weierstrass 定理，函数 g 在 S 上达到最大值与最小值. 于是，有 $\boldsymbol{x},\boldsymbol{y}\in\mathscr{S}$使得 $\|\boldsymbol{x}\|_2=\|\boldsymbol{y}\|_2=1$，并且，

$$\|\boldsymbol{x}\|\leqslant\|\boldsymbol{v}\|\leqslant\|\boldsymbol{y}\|,\qquad \forall\,\boldsymbol{v}\in\mathscr{S},\quad \|\boldsymbol{v}\|_2=1.$$

现对 $\boldsymbol{0}\neq\boldsymbol{u}\in\mathscr{S}$，有 $\|\boldsymbol{u}/\|\boldsymbol{u}\|_2\|_2=1$，因而

$$\|\boldsymbol{x}\|\leqslant\frac{\|\boldsymbol{u}\|}{\|\boldsymbol{u}\|_2}\leqslant\|\boldsymbol{y}\|.$$

取 $c=\|\boldsymbol{x}\|$ 与 $d=\|\boldsymbol{y}\|$，则有 $c,d>0$ 与

$$c\|\boldsymbol{u}\|_2\leqslant\|\boldsymbol{u}\|\leqslant d\|\boldsymbol{u}\|_2,\qquad \forall\,\boldsymbol{u}\in\mathscr{S}.$$ □

由定理 3 可直接推出如下重要结果.

推论 4　设$\{\boldsymbol{x}^{(k)}\}$为有限维线性空间 $\mathscr{S}$内的序列，且 $\boldsymbol{x}\in\mathscr{S}$. 若 $\|\cdot\|_\alpha$ 与 $\|\cdot\|_\beta$为 $\mathscr{S}$ 上任意两个向量范数，则$\lim\limits_{k\to\infty}\|\boldsymbol{x}^{(k)}-\boldsymbol{x}\|_\alpha=0$ 当且仅当$\lim\limits_{k\to\infty}\|\boldsymbol{x}^{(k)}-\boldsymbol{x}\|_\beta=0$.

今后，我们用 $\boldsymbol{x}^{(k)}\to\boldsymbol{x}\ (k\to\infty)$或$\lim\limits_{k\to\infty}\boldsymbol{x}^{(k)}=\boldsymbol{x}$ 表示 $\mathscr{S}$内向量序列$\{\boldsymbol{x}^{(k)}\}$(关于 $\mathscr{S}$ 上任意向量范数 $\|\cdot\|$)收敛于 $\boldsymbol{x}\in\mathscr{S}$. 定理 3 或推论 4 表明由不同向量范数诱导出 $\mathscr{S}$ 上不同的拓扑是彼此等价的. 但是，对于应用来说，在 $\mathscr{S}$上选择合适的向量范数却是十分重要的.

例 5　对于$\mathbb{C}^n$ 上的与 n 维线性空间 $\mathscr{S}$上的三种常用向量范数(见 2.1.1 小节(4)式与(6)式)，有如下的特殊等价不等式(3)：

$$\begin{aligned}&\|\boldsymbol{x}\|_\infty\leqslant\|\boldsymbol{x}\|_1\leqslant n\|\boldsymbol{x}\|_\infty,\\&\frac{1}{\sqrt{n}}\|\boldsymbol{x}\|_1\leqslant\|\boldsymbol{x}\|_2\leqslant\|\boldsymbol{x}\|_1,\\&\frac{1}{\sqrt{n}}\|\boldsymbol{x}\|_2\leqslant\|\boldsymbol{x}\|_\infty\leqslant\|\boldsymbol{x}\|_2.\end{aligned}\tag{5}$$

定理 3 另一个重要推论为 $\mathscr{S}$中每一个向量范数 $\|\cdot\|$ 的单位闭球$\{\boldsymbol{x}\in\mathscr{S}:\ \|\boldsymbol{x}\|\leqslant1\}$与单位球面$\{\boldsymbol{x}\in\mathscr{S}:\ \|\boldsymbol{x}\|=1\}$为 $\mathscr{S}$内紧集. 这个事实蕴涵定义在任一向量范数的单位闭球上的复值连续函数为有界的，且若此函数还为实值的，则它在此单位闭球上达到最大值与最小值.

推论 6　设 $\|\cdot\|$ 与 $\|\cdot\|_\alpha$ 均为有限维空间 $\mathscr{S}$上的向量范数. 则赋范空间$(\mathscr{S},\|\cdot\|)$内的单位闭球$\{\boldsymbol{x}\in\mathscr{S}:\ \|\boldsymbol{x}\|\leqslant1\}$与单位球面$\{\boldsymbol{x}\in\mathscr{S}:\ \|\boldsymbol{x}\|=1\}$皆为赋范空间$(\mathscr{S},\|\cdot\|_\alpha$内紧集.

证明　按定理 3，存在正数 $d>0$ 使得

$$\|\boldsymbol{x}\|_\alpha\leqslant d\|\boldsymbol{x}\|,\qquad \forall\,\boldsymbol{x}\in\mathscr{S},\text{这里，}\|\cdot\|_\alpha\ \text{为}\ \mathscr{S}\text{上给定的向量范数。}$$

于是，集合$\{x\in\mathscr{S}: \|x\|\leqslant 1\}$包含在半径为$d$，中心在原点的$(\mathscr{S}, \|\cdot\|_\alpha)$闭球$\{x\in\mathscr{S}: \|x\|_\alpha\leqslant d\}$内，因而它为有界的. 集合$\{x\in\mathscr{S}: \|x\|\leqslant 1\}$与$\{x\in\mathscr{S}: \|x\|=1\}$均为$(\mathscr{S}, \|\cdot\|_\alpha)$内的闭集，因为$\|\cdot\|$为$(\mathscr{S}, \|\cdot\|_\alpha)$内的连续函数. 再应用事实：在$\mathscr{S}$内有界闭集是紧的，因为$\mathscr{S}$同构于$\mathbb{C}^n$或$\mathbb{R}^n$，对于后者，有界闭性等同于紧性. □

现在考虑向量范数的**几何性质**. 向量范数的主要几何特性在于它的单位闭球.

令$\bar{B}_{\|\cdot\|}$为n维赋范空间$(\mathscr{S}, \|\cdot\|)$的单位闭球，则按推论 6，它是该空间的**紧集**.

$\bar{B}_{\|\cdot\|}$为空间$(\mathscr{S}, \|\cdot\|)$内的**凸集**，这里因为若$\|x\|\leqslant 1$，$\|y\|\leqslant 1$与$\alpha\in[0,1]$，则

$$\|\alpha x+(1-\alpha)y\|\leqslant\alpha\|x\|+(1-\alpha)\|y\|\leqslant\alpha+(1-\alpha)=1,$$

因此有$\alpha x+(1-\alpha)y\in\bar{B}_{\|\cdot\|}$.

$\bar{B}_{\|\cdot\|}$为空间$(\mathscr{S}, \|\cdot\|)$内的**均衡集**，即若$x\in\bar{B}_{\|\cdot\|}$，则对一切满足$|\alpha|=1$的$\alpha\in\mathbb{C}$，必有$\alpha x\in\bar{B}_{\|\cdot\|}$. 这个性质为向量范数齐次性的直接推论.

此外，$\bar{B}_{\|\cdot\|}$显然以$x=\mathbf{0}$为其内点.

通常，将空间$(\mathscr{S}, \|\cdot\|)$内的包含内点的紧凸集称为**凸体**(convex body). 从刚才讨论看出：有限维赋范空间$(\mathscr{S}, \|\cdot\|)$内的单位闭球$\bar{B}_{\|\cdot\|}$为含有内点$\mathbf{0}$的**均衡凸体**. 从下一定理知道，这个关于单位闭球的必要条件也是充分的，因而它可以作为向量范数的一个特征.

定理 7 设$\mathscr{S}$为有限维空间. 则$\mathscr{S}$内集合B为$\mathscr{S}$上某向量范数的单位闭球的充分与必要条件是B为有内点$\mathbf{0}$的均衡凸体.

证明 按前面讨论只要证明条件的充分性，即要证：若B为有内点$\mathbf{0}$的均衡凸体，则可以确定出$\mathscr{S}$上的某个向量范数$\|\cdot\|$，使得B为空间$(\mathscr{S}, \|\cdot\|)$的单位闭球. 任取$\mathbf{0}\neq x\in\mathscr{S}$，因为原点为$B$的内点，所以必存在$\nu>0$使得$x\in\nu B$. 定义实值函数$\|\cdot\|$如下：

$$\|x\|=\begin{cases}0, & 若\ x=\mathbf{0},\\ \min\{\nu:\nu>0, x\in\nu B\}, & 若\ x\neq\mathbf{0}.\end{cases}\tag{6}$$

此函数$\|\cdot\|$对所有$x\in\mathscr{S}$有定义，因为B为紧集且有内点$\mathbf{0}$. 它满足向量范数条件(1)与(4)是明显的. 应用B的均衡性假定容易看出，$\|\cdot\|$满足齐次性条件(2). 对于三角不等式，证明如下. 设$x\in\nu B$与$y\in\mu B$，由B的凸性则得$x+y\in\nu B+\mu B=(\nu+\mu)B$，因此，$\|x+y\|\leqslant\nu+\mu$. 令$\nu=\|x\|$与$\mu=\|y\|$，就证明了范数条件(3). 对于$\mathscr{S}$上这个向量范数$\|\cdot\|$，容易验证：$B$为它的单位闭球. □

上述定理表明，$\mathscr{S}$上向量范数在几何上可以借助于$\mathscr{S}$内某些均衡凸体来定义，二者之间有着一一对应关系.

习题 2.1.2

1. 试证:任一有限维赋范空间内的收敛向量序列必定为 Cauchy 序列,反之亦然.

2. 验明例 5 中的三个不等式.

3. 设 f_1 与 f_2 为有限维空间 $\mathscr{S}$ 上两个实值函数,$\mathscr{B}=\{\boldsymbol{b}_1,\cdots,\boldsymbol{b}_n\}$ 为 $\mathscr{S}$ 的基. 试证:若 $f_i(i=1,2)$ 为:

(1) 正的:$f_i(\boldsymbol{x})\geqslant 0, \forall \boldsymbol{x}\in\mathscr{S}, f_i(\boldsymbol{x})=0$ 当且仅当 $\boldsymbol{x}=\boldsymbol{0}$;

(2) 齐次的:$f_i(\alpha\boldsymbol{x})=|\alpha|f_i(\boldsymbol{x}), \forall \alpha\in\boldsymbol{F}$ 与 $\forall \boldsymbol{x}\in\mathscr{S}$;

(3) 连续的:$f_i(\boldsymbol{x}(\boldsymbol{\mu}))$ 在 $\boldsymbol{F}^n$ 上连续,其中

$$\boldsymbol{\mu}=(\mu_1,\cdots,\mu_n)^{\mathrm{T}}\in\boldsymbol{F}^n \text{ 与 } \boldsymbol{x}(\boldsymbol{\mu})=\sum_{j=1}^n \mu_j\boldsymbol{b}_j,$$

则存在有限正数 c 与 d 使得

$$cf_1(\boldsymbol{x})\leqslant f_2(\boldsymbol{x})\leqslant df_1(\boldsymbol{x}), \forall \boldsymbol{x}\in\mathscr{S}.$$

4. 设 $\mathscr{S}$ 为有限维线性空间. 则满足题 3 中条件(1),(2)与(3)的非负实值函数 $f:\mathscr{S}\to\mathbb{R}$ 叫作 $\mathscr{S}$ 上一个**准范数**(pre-norm). 试证:若 $\|\cdot\|_{\alpha_1},\cdots,\|\cdot\|_{\alpha_k}$ 为 $\mathscr{S}$ 上向量范数,则

$$f(\boldsymbol{x})=[\|\boldsymbol{x}\|_{\alpha_1}\cdots\|\boldsymbol{x}\|_{\alpha_k}]^{1/k} \quad \text{与} \quad h(\boldsymbol{x})=\min_{1\leqslant i\leqslant k}\|\boldsymbol{x}\|_{\alpha_i}$$

为 $\mathscr{S}$ 上的准范数,但不必为向量范数.

5. 设 $C[0,1]$ 表示定义在区间$[0,1]$上所有实值或复值连续函数的线性空间,它是无穷维的. 试用如下实例证明此空间上的 L_1 范数 $\|f\|_1=\int_0^1|f(t)|^2\mathrm{d}t$ 的收敛性不蕴涵关于 L_2 范数 $\|f\|_2=\left[\int_0^1|f(t)|^2\mathrm{d}t\right]^{1/2}$ 与 L_∞ 范数 $\|f\|_\infty=\max\{|f(t)|:t\in[0,1]\}$ 的收敛性:考虑 $C[0,1]$ 上的函数列$\{f_k\}$,

$$\begin{aligned}
&f_k(t)=0, && \text{如果 } 0\leqslant t\leqslant 1/k\\
&f_k(t)=2(k^{3/2}t-k^{1/2}), && \text{如果 } 1/k\leqslant t\leqslant 3/(2k)\\
&f_k(t)=2(-k^{3/2}t+2k^{1/2}), && \text{如果 } 3/(2k)\leqslant t\leqslant 2/k\\
&f_k(t)=0, && \text{如果 } 2/k\leqslant t\leqslant 1\\
&k=2,3,4,\cdots.
\end{aligned}$$

则 $\|f_k\|_1\to 0(k\to\infty)$, $\|f_k\|_2=1/\sqrt{3}, k=2,3,\cdots$, $\|f_k\|_\infty\to\infty(k\to\infty)$.

6. 设$(\mathscr{S},\|\cdot\|)$为 n 维赋范空间,X 与 Y 为它的不空子集,X 为有界闭集,而 Y 为闭集. 试证:存在 $\boldsymbol{x}_0\in X$ 与 $\boldsymbol{y}_0\in Y$ 使得

$$\|\boldsymbol{x}_0-\boldsymbol{y}_0\|=\mathrm{dist}(X,Y)\equiv\min_{\boldsymbol{x}\in X,\boldsymbol{y}\in Y}\|\boldsymbol{x}-\boldsymbol{y}\|,$$

式中 $\mathrm{dist}(X,Y)$ 表示集合 X 与 Y 之间的距离.

7. 用 $\mathrm{dist}(\boldsymbol{x},S)=\inf_{\boldsymbol{y}\in S}\|\boldsymbol{x}-\boldsymbol{y}\|$ 表示 $\boldsymbol{x}$ **到赋范空间**$(\mathscr{S},\|\cdot\|)$**子集 S 的距离**. 试证:若 S 为 $\mathscr{S}$ 的不空闭集,且 $\boldsymbol{x}\notin S$,则存在 $\boldsymbol{y}_0\in S$ 使得 $\mathrm{dist}(\boldsymbol{x},S)=\|\boldsymbol{x}-\boldsymbol{y}_0\|>0$.

8. 设 $B\subset\mathbb{C}^n$ 为凸体,且 $\boldsymbol{0}$ 为 B 的内点. 如下定义的函数 $\varphi^B:\mathbb{C}^n\to\mathbb{R}$ 叫作 **B 的度规函数**(gauge function of B):

$$\varphi^B(\boldsymbol{x})=\inf\{\lambda>0:\boldsymbol{x}\in\lambda B\}.$$

试证：

(1) φ^B 为**度规函数**，即 φ^B 在$\mathbb{C}^n$ 上连续；$\varphi^B(\boldsymbol{x})\geqslant 0$，$\forall \boldsymbol{x}\in\mathbb{C}^n$ 且仅当 $\boldsymbol{x}=\boldsymbol{0}$ 时 $\varphi^B(\boldsymbol{x})=0$；$\varphi^B(a\boldsymbol{x})=a\varphi^B(\boldsymbol{x})$，$\forall a\geqslant 0$ 与 $\forall \boldsymbol{x}\in\mathbb{C}^n$；$\varphi^B(\boldsymbol{x}+\boldsymbol{y})\leqslant\varphi^B(\boldsymbol{x})+\varphi^B(\boldsymbol{y})$，$\forall \boldsymbol{x},\boldsymbol{y}\in\mathbb{C}^n$.

(2) $B=\{\boldsymbol{x}\in\mathbb{C}^n:\varphi^B(\boldsymbol{x})\leqslant 1\}$.

(3) B 的所有内点组成的集合 int$B$ 恰是$\{\boldsymbol{x}\in\mathbb{C}^n:\varphi^B(\boldsymbol{x})<1\}$.

反之，若 $\varphi:\mathbb{C}^n\to\mathbb{R}$ 为任一度规函数(见本题(1)中定义)，则 φ 为下列凸体 B 的度规函数 φ^B：

$$B=\{\boldsymbol{x}\in\mathbb{C}^n:\varphi(\boldsymbol{x})\leqslant 1\},$$

这里，$\boldsymbol{0}\in\text{int}B$.

9. 设 B 为$\mathbb{C}^n$ 的凸体，$\boldsymbol{0}\in\text{int}B$. 试证 B 为均衡的当且仅当 B 的度规函数 φ^B 满足下列齐次性条件：

$$\varphi^B(a\boldsymbol{x})=|a|\varphi^B(\boldsymbol{x}),\quad \forall a\in\mathbb{C},\boldsymbol{x}\in\mathbb{C}^n.$$

10. 设 $\|\cdot\|_\alpha$ 与 $\|\cdot\|_\beta$ 均为 $\boldsymbol{F}^n$ 上向量范数. 试证：$\|\boldsymbol{x}\|_\alpha\leqslant\|\boldsymbol{x}\|_\beta$，$\forall \boldsymbol{x}\in\boldsymbol{F}^n$ 当且仅当 $B_{\|\cdot\|_\beta}\subset B_{\|\cdot\|_\alpha}$.

2.2 矩阵范数

2.2.1 广义矩阵范数

假设矩阵 $A\in\boldsymbol{M}_{m,n}(\boldsymbol{F})$，这里 $\boldsymbol{F}=\mathbb{C}$ 或$\mathbb{R}$，我们来考虑它的度量，即范数. 因为 $\boldsymbol{M}_{m,n}(\boldsymbol{F})$本身为 $\boldsymbol{F}$ 上 mn 维线性空间，所以任一 $m\times n$ 矩阵 A 可以用这个 mn 维空间上的某个向量范数来度量，并且，**2.1** 节中关于一般有限维的线性空间向量范数的结果对此特殊情形仍然有效. 通常，将定义在空间 $\boldsymbol{M}_{m,n}(\boldsymbol{F})$上的任一个向量范数都称为**广义矩阵范数**(以区别后面定义的矩阵范数)，确切地说有

定义 1 $\boldsymbol{M}_{m,n}(\boldsymbol{F})$上的**广义矩阵范数**(generalized matrix norm)是一个实值函数 $\|\cdot\|:\boldsymbol{M}_{m,n}(\boldsymbol{F})\to\mathbb{R}$，它满足 2.1.1 小节定义 1 中条件(1)～(4)，其中，$\mathscr{S}$由 $\boldsymbol{M}_{m,n}(\boldsymbol{F})$代替，$\boldsymbol{x}$ 与 $\boldsymbol{y}$ 由 $\boldsymbol{M}_{m,n}(\boldsymbol{F})$中矩阵 A 与 B 代替.

这样一来，对应于 2.1.1 小节(4)式中几种向量范数，我们有下列的 $\boldsymbol{M}_{m,n}(\boldsymbol{F})$上的广义矩阵范数：对 $A=(a_{ij})\in\boldsymbol{M}_{m,n}(\boldsymbol{F})$，

$$\|A\|'_\infty=\max_{i,j}|a_{ij}|,$$

$$\|A\|'_1=\sum_{i,j}|a_{ij}|,\qquad \|A\|''_1=\frac{1}{n}\sum_{i=1,j=1}^{m,n}|a_{ij}|. \tag{1}$$

$$\|A\|'_2=\Big(\sum_{i,j}|a_{ij}|^2\Big)^{1/2},$$

$$\|A\|'_p=\Big(\sum_{i,j}|a_{ij}|^p\Big)^{1/p},\quad p\geqslant 1,$$

其中，广义矩阵范数 $\|A\|'_2$ 叫作 A 的 **Frobenius 范数**或者**欧氏范数**，通常记为 $\|A\|_F$. 若 $A=[\boldsymbol{a}_1\ \boldsymbol{a}_2\cdots\ \boldsymbol{a}_n]\in\boldsymbol{M}_{m,n}(\boldsymbol{F})$，这里 $\boldsymbol{a}_j\in\boldsymbol{F}^m$ 为 A 的第 j 个列向量，则按定

义有

$$\| A \|_F^2 = \| \boldsymbol{a}_1 \|_2^2 + \cdots + \| \boldsymbol{a}_n \|_2^2 = \mathrm{tr}(A^* A) = \mathrm{tr}(AA^*). \tag{2}$$

由于$\mathbb{C}^m$上的l_2范数为酉不变的(见习题2.1.1题2),故只要$U \in \boldsymbol{M}_m(\boldsymbol{F})$为酉矩阵便有

$$\| UA \|_F^2 = \| U\boldsymbol{a}_1 \|_2^2 + \cdots + \| U\boldsymbol{a}_n \|_2^2 = \| \boldsymbol{a}_1 \|_2^2 + \cdots + \| \boldsymbol{a}_n \|_2^2 = \| A \|_F^2,$$

又由于$\| B^* \|_F = \| B \|_F$, $\forall B \in \boldsymbol{M}_{m,n}(\boldsymbol{F})$,故只要$U \in \boldsymbol{M}_m(\boldsymbol{F})$与$V \in \boldsymbol{M}_n(\boldsymbol{F})$为酉矩阵便有

$$\| UAV \|_F = \| AV \|_F = \| V^* A^* \|_F = \| A^* \|_F = \| A \|_F. \tag{3}$$

这表明,$\boldsymbol{M}_{m,n}(\boldsymbol{F})$上的 Frobenius 范数为一种**酉不变的广义矩阵范数**.

应用有限维空间上的向量范数等价性定理(2.1.2小节定理3)可以得出:$\boldsymbol{M}_{m,n}(\boldsymbol{F})$上任意两个广义矩阵范数是等价的. 设$A^{(k)} = (a_{ij}^{(k)}) \in \boldsymbol{M}_{m,n}(\boldsymbol{F})$, $k=1, 2, \cdots$. 如果

$$\lim_{k\to\infty} a_{ij}^{(k)} = a_{ij}, i = 1, \cdots, m; j = 1, \cdots, n, \tag{4}$$

则称矩阵序列$\{A^{(k)}\}$**收敛**于$A=(a_{ij}) \in \boldsymbol{M}_{m,n}(\boldsymbol{F})$,或它有**极限**$A$,记作

$$A^{(k)} \to A (k \to \infty) \quad 或者 \quad \lim_{k\to\infty} A^{(k)} = A.$$

这相当于:$\lim\limits_{k\to\infty} \| A^{(k)} - A \|'_\infty = 0$. 因此,对应于2.1.2小节推论4,我们有:若$\{A^{(k)}\}$为$\boldsymbol{M}_{m,n}(\boldsymbol{F})$内矩阵序列,$\|\cdot\|$为$\boldsymbol{M}_{m,n}(\boldsymbol{F})$上任一个广义矩阵范数,则

$$A^{(k)} \to A (k \to \infty) \Leftrightarrow \lim_{k\to\infty} \| A^{(k)} - A \| = 0. \tag{5}$$

我们经常遇到矩阵之间或矩阵与向量之间的乘法运算,因而自然希望:若$A \in \boldsymbol{M}_{m,n}(\boldsymbol{F})$, $B \in \boldsymbol{M}_{n,l}(\boldsymbol{F})$,则对同一广义矩阵范数$\|\cdot\|$(但定义在不同空间上)有

$$\| AB \| \leqslant \| A \| \ \| B \|. \tag{6}$$

满足条件(6)的广义矩阵范数$\|\cdot\|$称为**具备相容性条件**的. 取B为$\boldsymbol{x} \in \boldsymbol{F}^n$,则(6)式变为

$$\| A\boldsymbol{x} \| \leqslant \| A \| \ \| \boldsymbol{x} \|. \tag{7}$$

我们指出,不是所有广义矩阵范数都具备相容性条件(6),例如,前面讲过的$\|\cdot\|'_\infty$与$\|\cdot\|''_1$不满足条件(6). 为此考虑

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad 与 \quad B = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix},$$

我们有$\| A \|'_\infty = \| B \|'_\infty = 1$,而$\| AB \|'_\infty = 2$; $\| A \|''_1 = \| B \|''_1 = 3/2$,而$\| AB \|''_1 = 5/2$. 但是,Frobenius 范数$\|\cdot\|_F$具备相容性条件,即有

$$\| AB \|_F \leqslant \| A \|_F \| B \|_F, \qquad \forall A \in \boldsymbol{M}_{m,n}(\boldsymbol{F}), B \in \boldsymbol{M}_{n,l}(\boldsymbol{F}). \tag{8}$$

事实上,若用$\rho(A^* A)$表示$A^* A$的谱半径,则由(2)式与习题5结果推出,

$$\| AB \|_F^2 = \mathrm{tr}(B^* A^* AB) = \mathrm{tr}(A^* ABB^*)$$
$$\leqslant \rho(A^* A) \mathrm{tr}(B^* B)$$

$$\leqslant \operatorname{tr}(A^* A)\operatorname{tr}(B^* B) = \|A\|_F^2 \|B\|_F^2.$$

不同的广义矩阵范数之间也可能具有“相容性”.

定义 2 设 $\|\cdot\|_\alpha$，$\|\cdot\|_\beta$ 与 $\|\cdot\|_\gamma$ 分别为 $\boldsymbol{M}_{m,n}(\boldsymbol{F})$ 上，$\boldsymbol{M}_{n,l}(\boldsymbol{F})$ 上与 $\boldsymbol{M}_{m,l}(\boldsymbol{F})$ 上的广义矩阵范数. 若

$$\|AB\|_\gamma \leqslant \|A\|_\alpha \|B\|_\beta \qquad \forall A \in \boldsymbol{M}_{m,n}(\boldsymbol{F}), B \in \boldsymbol{M}_{n,l}(\boldsymbol{F}),$$

则称 $\|\cdot\|_\alpha$，$\|\cdot\|_\beta$ 与 $\|\cdot\|_\gamma$ 是**相容的**.

这个定义包含广义矩阵范数与向量范数的相容性定义. 满足这种相容性的有

$$\|Ax\|_2 \leqslant \|A\|_F \|\boldsymbol{x}\|_2, \qquad \forall A \in \boldsymbol{M}_{m,n}(\boldsymbol{F}), \boldsymbol{x} \in \boldsymbol{F}^n. \tag{9}$$

这是(8)式的特殊情形.

现在考虑 $\boldsymbol{M}_{m,n}(\boldsymbol{F})$ 上的一种特殊的广义矩阵范数，它由 $\boldsymbol{F}^m$ 与 $\boldsymbol{F}^n$ 上的向量范数诱导出来. 这种广义矩阵范数与诱导它的向量范数是相容的.

定义 3 设 $\|\cdot\|_\delta$ 与 $\|\cdot\|_\tau$ 分别为 $\boldsymbol{F}^n$ 与 $\boldsymbol{F}^m$ 上的向量范数. 由下式定义的非负实值函数 $\|\cdot\|_{\tau\delta}:\boldsymbol{M}_{m,n}(\boldsymbol{F})\to\mathbb{R}$：

$$\|A\|_{\tau\delta} = \sup_{0\neq \boldsymbol{x}\in \boldsymbol{F}^n} \frac{\|A\boldsymbol{x}\|_\tau}{\|\boldsymbol{x}\|_\delta} = \sup_{\|\boldsymbol{x}\|_\delta = 1} \|A\boldsymbol{x}\|_\tau = \sup_{\|\boldsymbol{x}\|_\delta \leqslant 1} \|A\boldsymbol{x}\|_\tau \tag{10}$$

叫作**向量范数 $\|\cdot\|_\delta$ 与 $\|\cdot\|_\tau$ 导出的广义矩阵范数或算子范数.**

容易验证(10)式中不同表达式的等式成立，并且，由于赋范空间($\boldsymbol{F}^n$，$\|\cdot\|_\delta$)的单位闭球与单位球面皆为紧集，而 $\|A\boldsymbol{x}\|_\tau$ 为 $\boldsymbol{x}$ 的实值连续函数，故(10)式中的“sup”均可以代换为“max”.

显然，(10)式等价于如下两个条件：

(1) $\|A\boldsymbol{x}\|_\tau \leqslant \|A\|_{\tau\delta} \|\boldsymbol{x}\|_\delta$，$\forall \boldsymbol{x}\in \boldsymbol{F}^n$；

(2) 存在 $\boldsymbol{0}\neq \boldsymbol{x}_0 \in \boldsymbol{F}^n$，使得

$$\|A\boldsymbol{x}_0\|_\tau = \|A\|_{\tau\delta} \|\boldsymbol{x}_0\|_\delta.$$

我们指出，由(10)式定义的 $\boldsymbol{M}_{m,n}(\boldsymbol{F})$ 上实值函数 $\|\cdot\|_{\tau\delta}$ 确实为 $\boldsymbol{M}_{m,n}(\boldsymbol{F})$ 上的广义矩阵范数，因为容易直接验证，它满足定义 1 中所有条件. 同时，关于这样导出的广义矩阵范数的相容性，有如下重要结果：

$$\|AB\|_{\tau\omega} \leqslant \|A\|_{\tau\delta} \|B\|_{\delta\omega}, \qquad \forall A \in \boldsymbol{M}_{m,n}(\boldsymbol{F}), B \in \boldsymbol{M}_{n,l}(\boldsymbol{F}), \tag{11}$$

式中，$\|\cdot\|_\tau$，$\|\cdot\|_\delta$ 与 $\|\cdot\|_\omega$ 分别为 $\boldsymbol{F}^m$，$\boldsymbol{F}^n$ 与 $\boldsymbol{F}^l$ 上的向量范数. 事实上，对任意 $\boldsymbol{x}\in \boldsymbol{F}^l$ 有 $\|AB\boldsymbol{x}\|_\tau \leqslant \|A\|_{\tau\delta} \|B\boldsymbol{x}\|_\delta \leqslant \|A\|_{\delta\omega} \|B\|_{\delta\omega} \|\boldsymbol{x}\|_\omega$.

一般地说，已知向量范数 $\|\cdot\|_\delta$ 与 $\|\cdot\|_\tau$，很难求得计算相应导出的广义矩阵范数的明显公式，但也有例外情形. 下面结果提供三种基本导出的广义矩阵范数公式.

定理 4 设 $A=(a_{ij})\in \boldsymbol{M}_{m,n}(\boldsymbol{F})$. 则：

(1) 若 $\|\cdot\|_1$ 为 $\boldsymbol{F}^n$ 与 $\boldsymbol{F}^m$ 上的 l_1 范数，那么它们导出的广义矩阵范数为

$$\|A\|_1 = \max_j \sum_i |a_{ij}|. \tag{12}$$

(2) 若 $\|\cdot\|_\infty$ 为 $\boldsymbol{F}^n$ 与 $\boldsymbol{F}^m$ 上的 l_∞ 范数，那么它们导出的广义矩阵范数为

$$\| A \|_\infty = \max_i \sum_j | a_{ij} |. \tag{13}$$

(3) 若 $\|\cdot\|_2$ 为 $\boldsymbol{F}^n$ 与 $\boldsymbol{F}^m$ 上的 l_2 范数，那么它们导出的广义矩阵范数为

$$\| A \|_2 = (\rho(A^* A))^{1/2} \qquad \text{（谱范数）}. \tag{14}$$

证明 (1) 设 $\boldsymbol{x}=(x_1,\cdots,x_n)^{\mathrm{T}}\in \boldsymbol{F}^n$. 则

$$\begin{aligned}\| A\boldsymbol{x} \|_1 &= \sum_{i=1}^m \left| \sum_{j=1}^n a_{ij}x_j \right| \leqslant \sum_{j=1}^n \left(\sum_{i=1}^m | a_{ij} | \right) | x_j | \\ &\leqslant \left(\max_{1\leqslant j\leqslant n} \sum_{i=1}^m | a_{ij} | \right) \sum_{j=1}^n | x_j | \\ &= \left(\max_{1\leqslant j\leqslant n} \sum_{i=1}^m | a_{ij} | \right) \| \boldsymbol{x} \|_1. \end{aligned}\tag{15}$$

另一方面，若 k 满足

$$\max_j \sum_i | a_{ij} | = \sum_i | a_{ik} |,$$

置 $\boldsymbol{x}_0=\boldsymbol{e}_k$（$\boldsymbol{F}^n$ 的第 k 个单位坐标向量），则 $\boldsymbol{x}_0\neq\boldsymbol{0}$，并且，

$$\| A\boldsymbol{x}_0 \|_1 = \sum_i | a_{ik} | = \left(\max_j \sum_i | a_{ij} | \right) \| \boldsymbol{x}_0 \|_1.$$

因此，(12)式成立.

(2) 对任意 $\boldsymbol{x}\in \boldsymbol{F}^n$ 有

$$\begin{aligned}\| A\boldsymbol{x} \|_\infty &= \max_i \left| \sum_j a_{ij}x_j \right| \\ &\leqslant \max_i \sum_j | a_{ij} | \, | x_j | \leqslant \left(\max_i \sum_j | a_{ij} |\right) \max_j | x_j | \\ &= \left(\max_i \sum_j | a_{ij} |\right) \| \boldsymbol{x} \|_\infty. \end{aligned}$$

另一方面，若 k 满足

$$\max_i \sum_j | a_{ij} | = \sum_j | a_{kj} |,$$

置 $\boldsymbol{x}_0=(x_{01},\cdots,x_{0n})^{\mathrm{T}}$，其中，

$$x_{0j} = \begin{cases} | a_{kj} | / a_{kj}, & \text{如果 } a_{kj} \neq 0, \\ 1, & \text{如果 } a_{kj} = 0, \end{cases}$$

则 $\boldsymbol{x}_0\neq\boldsymbol{0}$，并且，

$$\begin{aligned}\| A\boldsymbol{x}_0 \|_\infty &= \max_i \left| \sum_j a_{ij}x_{0j} \right| \geqslant \left| \sum_j a_{kj}x_{0j} \right| = \sum_j | a_{kj} | \\ &= \left(\max_i \sum_j | a_{ij} |\right) \| \boldsymbol{x}_0 \|_\infty, \end{aligned}$$

因为 $|x_{0j}|=1, \forall j$. 但根据前面建立的公式，

$$\| Ax_0 \|_\infty \leqslant (\max_i \sum_j |a_{ij}|) \| x_0 \|_\infty,$$

因而上式变为等式.(13)式得证.

(3) 对任意 $x \in F^n$,按第一章 1.2.1 小节定理 6 后面说明,

$$\begin{aligned} \| Ax \|_2 &= (Ax, Ax)^{1/2} = (A^* Ax, x)^{1/2} \\ &\leqslant (\rho(A^* A))^{1/2} (x, x)^{1/2} = (\rho(A^* A))^{1/2} \| x \|_2, \end{aligned}$$

式中,$\rho(A^* A)$为 Hermite 矩阵 $A^* A$ 的谱半径.另一方面,若 $x_0 \neq 0$ 且 $A^* Ax_0 = \rho(A^* A)x_0$,则

$$\begin{aligned} \| Ax_0 \|_2 &= (Ax_0, Ax_0)^{1/2} = (A^* Ax_0, x_0)^{1/2} \\ &= (\rho(A^* A)x_0, x_0)^{1/2} = (\rho(A^* A))^{1/2} \| x_0 \|_2. \end{aligned}$$

因此,(14)式成立. □

我们指出,在许多问题中经常使用的 $M_{m,n}(F)$上广义矩阵范数 $\|\cdot\|_F$(对$m>1$ 或 $n>1$)不是 F^n 与 F^m 上向量范数导出的广义矩阵范数,但是,对 F^n 与 F^m 上欧氏范数与 $M_{m,n}(F)$上 Frobenius 范数,我们有相容性不等式(9).

习题 2.2.1

1. 试证:虽然(1)式中 $\|\cdot\|'_\infty$ 不满足条件(6),但当 $m=n$ 时,$|||A||| = n \| A \|'_\infty$, $\forall A \in M_n(F)$,为 $M_n(F)$上满足$|||AB||| \leqslant |||A|||\ |||B|||$, $\forall A, B \in M_n(F)$的广义矩阵范数.

2. 证明:由(14)式定义的 $M_{m,n}(F)$上广义矩阵范数为酉不变的,但是由(12)式,(13)式定义的不是酉不变的.

3. 设 $\|\cdot\|$ 为 $M_{m,n}(F)$上广义矩阵范数,使得对 $M_{m,n}(F)$中任意秩为 1 的矩阵 A 有 $\| A \| = \| A \|_2$.试证:若 $A = xy^*$, $x \in F^m$, $y \in F^n$, $x \neq 0$, $y \neq 0$,则

$$\| A \| = \| x \|_2 \| y \|_2.$$

4. 设 $A \in M_{m,n}(F)$的非零奇异值为 $\sigma_1, \sigma_2, \cdots, \sigma_r$,试证:

$$\| A \|_2 = \max_i \sigma_i \text{ 与 } \| A \|_F = \left(\sum_{j=1}^r \sigma_j^2 \right)^{1/2}.$$

5. 在证明(8)式时,应用事实:若 P 与 Q 为同阶的正半定矩阵,则 $\mathrm{tr}(PQ) \leqslant \mathrm{tr}P\rho(Q)$与 $\mathrm{tr}(PQ) \leqslant \rho(P)\mathrm{tr}Q$.试证之.并且应用这个事实证明:若 A 与 B 使得乘积AB 存在,则有

$$\| AB \|_F \leqslant \| A \|_F \| B \|_2,$$

$$\| AB \|_F \leqslant \| A \|_2 \| B \|_F.$$

6. 给出 $M_n(F)$上一个广义矩阵范数 $\|\cdot\|$,使得

$$\| I_n \| < 1.$$

2.2.2 矩阵范数

在 2.2.1 小节知识基础上,本小节讨论矩阵范数,它限定在方阵情形.

定义 1 若 $M_n(F)$上的广义矩阵范数 $\|\cdot\|$ 满足如下**次可乘条件**或**相容性条件**:

$$\| AB \| \leqslant \| A \| \ \| B \|, \forall A,B \in \boldsymbol{M}_n(\boldsymbol{F}), \tag{1}$$

则称 $\|\cdot\|$ 为 $\boldsymbol{M}_n(\boldsymbol{F})$上的**矩阵范数**.

由 2.2.1 小节(8)式知道,$\boldsymbol{M}_n(\boldsymbol{F})$上的 Frobenius 范数为矩阵范数,但 $\boldsymbol{M}_n(\boldsymbol{F})$上的 $\|\cdot\|'_\infty$ 与 $\|\cdot\|''_1$ 不是矩阵范数,因为它们都不满足条件(1). 此外,在 2.2.1 小节的定义 3 中,若限定 $m=n$,且 $\|\cdot\|_\delta$ 与 $\|\cdot\|_\tau$ 取为 $\boldsymbol{F}^n$ 上同一向量范数,则由 2.2.1 小节(10)式确定的 $\boldsymbol{M}_n(\boldsymbol{F})$上的广义矩阵范数:

$$\| A \| = \sup_{0\neq x\in \boldsymbol{F}^n} \frac{\| A\boldsymbol{x} \|_\delta}{\| \boldsymbol{x} \|_\delta}, \forall A \in \boldsymbol{M}_n(\boldsymbol{F}) \tag{2}$$

也满足条件(1)(见 2.2.1 小节(11)式),因而它也为 $\boldsymbol{M}_n(\boldsymbol{F})$上的矩阵范数,通常称为**向量范数 $\|\cdot\|_\delta$ 导出的矩阵范数或算子范数**. 因此,根据 2.2.1 小节定理 4,对任意 $A\in\boldsymbol{M}_n(\boldsymbol{F})$,

$$\| A \|_1 = \max_{1\leqslant j\leqslant n}\sum_{i=1}^{n} | a_{ij} | \qquad (A \text{ 的}\textbf{列范数})$$

$$\| A \|_\infty = \max_{1\leqslant i\leqslant n}\sum_{j=1}^{n} | a_{ij} | \qquad (A \text{ 的}\textbf{行范数}),$$

$$\| A \|_2 = (\rho(A^* A))^{1/2} \qquad (A \text{ 的}\textbf{谱范数})$$

均为 $\boldsymbol{M}_n(\boldsymbol{F})$上矩阵范数,更确切地说,它们分别是 $\boldsymbol{F}^n$ 上向量范数 $\|\cdot\|_1$, $\|\cdot\|_\infty$ 与 $\|\cdot\|_2$(见 2.1.1 小节(4)式)导出的矩阵范数.

由(2)式看出,对任意导出的矩阵范数 $\|\cdot\|$,有

$$\| I \| = 1.$$

因此,当 $n>1$ 时,$\|\cdot\|_F$ 不是 $\boldsymbol{M}_n(F)$上导出的矩阵范数,因为 $\| I_n \|_F=\sqrt{n}$. 但根据矩阵范数定义,对 $\boldsymbol{M}_n(\boldsymbol{F})$上任意矩阵范数 $\|\cdot\|$,我们有(见习题 1(2))

$$\begin{aligned} &\| I_n \| \geqslant 1, \| A^k \| \leqslant \| A \|^k (k \text{ 为正整数});\\ &\| A^{-1} \| \geqslant \| I_n \| / \| A \| \geqslant \| A \|^{-1}, \text{若 } A \text{ 非奇异};\\ &\| A \| \geqslant 1, \text{若 } A \text{ 为非零幂等矩阵}. \end{aligned} \tag{3}$$

由下面定理看出,一种 $\boldsymbol{M}_n(\boldsymbol{F})$上的矩阵范数,通过相似变换可以引出另一种 $\boldsymbol{M}_n(\boldsymbol{F})$上的矩阵范数.

定理 2　若 $\|\cdot\|$ 为 $\boldsymbol{M}_n(\boldsymbol{F})$上矩阵范数,$S\in\boldsymbol{M}_n(\boldsymbol{F})$非奇异,则如下的 $\boldsymbol{M}_n(\boldsymbol{F})$上实值函数:

$$\| A \|_S = \| S^{-1}AS \|, \forall A \in \boldsymbol{M}_n(\boldsymbol{F}) \tag{4}$$

为 $\boldsymbol{M}_n(\boldsymbol{F})$上矩阵范数.

证明　直接验证便知,$\|\cdot\|_S$ 为 $\boldsymbol{M}_n(\boldsymbol{F})$上广义矩阵范数. 条件(1)证明如下:

$$\begin{aligned} \| AB \|_S &= \| S^{-1}ABS \| = \| (S^{-1}AS)(S^{-1}BS) \| \leqslant \| S^{-1}AS \| \ \| S^{-1}BS \| \\ &= \| A \|_S \| B \|_S. \end{aligned}$$

□

矩阵范数有一些重要的性质. 下一结果表明,任意矩阵范数必有与它相容的向

量范数(一个特例可见 2.2.1 小节(9)式).

定理 3 若 $\|\cdot\|$ 为 $\boldsymbol{M}_n(\boldsymbol{F})$ 上矩阵范数,则有 $\boldsymbol{F}^n$ 上向量范数 $\|\cdot\|_\alpha$ 满足

$$\|A\boldsymbol{x}\|_\alpha \leqslant \|A\| \ \|\boldsymbol{x}\|_\alpha, \qquad \forall A \in \boldsymbol{M}_n(\boldsymbol{F}), \boldsymbol{x} \in \boldsymbol{F}^n. \tag{5}$$

证明 对固定 $\boldsymbol{0}\neq\boldsymbol{a}\in\boldsymbol{F}^n$,定义

$$\|\boldsymbol{x}\|_\alpha = \|\boldsymbol{x}\boldsymbol{a}^{\mathrm{T}}\|, \qquad \forall \boldsymbol{x} \in \boldsymbol{F}^n.$$

则容易看出 $\|\cdot\|_\alpha$ 为 $\boldsymbol{F}^n$ 上向量范数. 此时由

$$\|A\boldsymbol{x}\|_\alpha = \|A\boldsymbol{x}\boldsymbol{a}^{\mathrm{T}}\| \leqslant \|A\| \ \|\boldsymbol{x}\boldsymbol{a}^{\mathrm{T}}\| = \|A\| \ \|\boldsymbol{x}\|_\alpha$$

推出相容关系(5). □

矩阵范数另一个重要性质为它给出矩阵谱半径的一个上界(见后面的定理 4). 我们指出,$\boldsymbol{M}_n(\mathbb{C})$ 上矩阵范数与矩阵谱半径之间有着密切的关系. 特别当 $A\in\boldsymbol{M}_n(\mathbb{C})$ 为 Hermite 矩阵时,A 的谱范数公式给出 $\|A\|_2=\rho(A)$. 但谱半径(作为 $\boldsymbol{M}_n(\mathbb{C})$ 上实值函数)本身不是 $\boldsymbol{M}_n(\mathbb{C})$ 上的矩阵范数(甚至不是广义矩阵范数)(见习题 12),并且,在某些情况下,$\rho(A)$ 与 A 的某些矩阵范数之间可能相差很大. 例如,对于二阶矩阵

$$A = \begin{bmatrix} 0 & 0 \\ \alpha & 0 \end{bmatrix}, \qquad \alpha \in \mathbb{C}, \tag{6}$$

若 $\alpha\neq 0$,则 $A\neq O$,且 $\|A\|_2=\|A\|_1=\|A\|_\infty=|\alpha|$,但是,$\rho(A)=0$, $\forall\alpha\in\mathbb{C}$.

定理 4 设 $A\in\boldsymbol{M}_n(\mathbb{C})$. 则对 $\boldsymbol{M}_n(\mathbb{C})$ 上任意矩阵范数 $\|\cdot\|$ 都有

$$\rho(A) \leqslant \|A\|. \tag{7}$$

证明 按定理 3,在 $\mathbb{C}^n$ 上有与 $\|\cdot\|$ 相容的向量范数 $\|\cdot\|_\alpha$. 现设 $\boldsymbol{0}\neq\boldsymbol{x}\in\mathbb{C}^n$ 为 A 对应于特征值 λ 的特征向量:$A\boldsymbol{x}=\lambda\boldsymbol{x}$. 则由 $\|\boldsymbol{x}\|_\alpha>0$ 与

$$|\lambda| \ \|\boldsymbol{x}\|_\alpha = \|\lambda\boldsymbol{x}\|_\alpha = \|A\boldsymbol{x}\|_\alpha \leqslant \|A\| \ \|\boldsymbol{x}\|_\alpha$$

推得 $|\lambda|\leqslant\|A\|$, $\forall\lambda\in\sigma(A)$,因而(7)式成立. □

另一方面,对固定的 $A\in\boldsymbol{M}_n(\mathbb{C})$,$A$ 的谱半径 $\rho(A)$ 为 $\boldsymbol{M}_n(\mathbb{C})$ 上 A 的所有矩阵范数值的下确界. 这个事实得自下列定理.

定理 5 设 $A\in\boldsymbol{M}_n(\mathbb{C})$ 且 $\varepsilon>0$ 为任意给定的正数. 则至少存在一种 $\boldsymbol{M}_n(\mathbb{C})$ 上的矩阵范数 $\|\cdot\|$,使得

$$\rho(A) \leqslant \|A\| \leqslant \rho(A) + \varepsilon. \tag{8}$$

证明 令 $D=\mathrm{diag}[1,\varepsilon,\varepsilon^2,\cdots,\varepsilon^{n-1}]$,且 $J=P^{-1}AP$ 为 A 的 Jordan 标准型. 则容易验证,$\hat{J}\equiv D^{-1}JD\in\boldsymbol{M}_n(\mathbb{C})$ 的元素除了用 ε 替换 J 中非对角线上每一个为 1 的元素以外,其余的均等于 J 的对应元素. 因此,按矩阵行范数 $\|\cdot\|_\infty$ 定义,

$$\|\hat{J}\|_\infty \leqslant \rho(A) + \varepsilon.$$

现记 $Q\equiv PD$,并定义 $\|B\|=\|Q^{-1}BQ\|_\infty$, $\forall B\in\boldsymbol{M}_n(\mathbb{C})$. 按照前面的定理 2,$\|\cdot\|$ 为 $\boldsymbol{M}_n(\mathbb{C})$ 上的矩阵范数,且有

$$\|A\| = \|Q^{-1}AQ\|_\infty = \|\hat{J}\|_\infty \leqslant \rho(A) + \varepsilon,$$

再引用(7)式便推出(8)式. □

这个定理表明,对任一固定的$A\in \boldsymbol{M}_n(\mathbb{C})$与给定的$\varepsilon>0$,总可以找到$\boldsymbol{M}_n(\mathbb{C})$上矩阵范数$\|\cdot\|$,使得

$$0\leqslant \|A\|-\rho(A)\leqslant\varepsilon, \tag{9}$$

亦即有

$$\rho(A)=\inf\{\|A\|:\|\cdot\| \text{ 为 } \boldsymbol{M}_n(\mathbb{C}) \text{ 上任意矩阵范数}\}. \tag{9'}$$

由前面讨论知道,对一般$A\in \boldsymbol{M}_n(\mathbb{C})$不是总能找到一种$\boldsymbol{M}_n(\mathbb{C})$上的矩阵范数$\|\cdot\|$,使得$\|A\|=\rho(A)$. 因此,从这个意义来讲,(9)式或(9′)式给出矩阵谱半径与矩阵范数值之间一般关系的最好结果.

应用定理4与定理5,我们可以讨论$A\in \boldsymbol{M}_n(\mathbb{C})$的乘幂序列$A,A^2,A^3,\cdots$的收敛性问题. 下面两个定理分别给出$\lim\limits_{k\to\infty}A^k=O$与$\lim\limits_{k\to\infty}A^k$存在的充分与必要条件. 今后要多次引用这些基本结论.

定理 6　设$A\in \boldsymbol{M}_n(\mathbb{C})$. 则$\lim\limits_{k\to\infty}A^k=O$当且仅当$\rho(A)<1$.

证明　若$A^k\to O$但$\rho(A)\geqslant 1$,那么,$\rho(A^k)=(\rho(A))^k\geqslant 1$,$k=1,2,\cdots$. 根据定理4,$\|A^k\|\geqslant\rho(A^k)\geqslant 1$对任一$\boldsymbol{M}_n(\mathbb{C})$上的矩阵范数$\|\cdot\|$与任意正整数$k$成立. 此显然与$A^k\to O$等价于$\lim\limits_{k\to\infty}\|A^k\|=0$的事实抵触. 反之,若$\rho(A)<1$,则按定理5,有$\boldsymbol{M}_n(\mathbb{C})$上某矩阵范数$\|\cdot\|$,使得$\|A\|<1$. 因此,$\|A^k\|\leqslant\|A\|^k\to 0(k\to\infty)$,此表明$A^k\to O(k\to\infty)$. □

定理 7　设$A\in \boldsymbol{M}_n(\mathbb{C})$. 则$\lim\limits_{k\to\infty}A^k$存在当且仅当下列三个条件均成立:

(1) $\rho(A)\leqslant 1$.

(2) 若$\rho(A)=1$,则模为1的A的所有特征值所对应的Jordan块均为一阶的.

(3) 若$\rho(A)=1$,则模为1的A的特征值等于1.

此外,当$k\to\infty$时,$\|A^k\|$有界当且仅当上述条件(1)与(2)成立,这里$\|\cdot\|$为$\boldsymbol{M}_n(\mathbb{C})$上任一矩阵范数.

证明　设$J=P^{-1}AP$为A的Jordan正规形式. 若$\lim\limits_{k\to\infty}A^k$存在,则$\lim\limits_{k\to\infty}J^k$存在,反之亦然,于是对任一$\lambda\in\sigma(A)$,$|\lambda^k|\leqslant 1$,此蕴涵$\rho(A)\leqslant 1$即条件(1). 今假定$\rho(A)=1$,且$\lambda\in\sigma(A)$有$|\lambda|=1$,这时与$\lambda$对应的$A$的Jordan块$J_\lambda$必定是一阶的,否则按第一章习题1.1.2题9,当$k\to\infty$时,$J_\lambda^k$(因而$A^k$)无界从而没有极限. 条件(2)得证. 此外,这时λ一定为1,否则J_λ^k的对角元素λ^k当$k\to\infty$时绕着复平面上单位圆周按同一方向转动,因而J^k(与A^k)没有极限. 同理可证:若$\|A^k\|$有界($k\to\infty$),则条件(1)与(2)亦成立.

反之,若$\rho(A)<1$,则由定理6知道,$A^k\to O$. 若$\rho(A)=1$,这时由于条件(2)与(3)成立,A必有如下的Jordan正规形式:

$$P^{-1}AP = \begin{bmatrix} A_1 & O \\ O & A_2 \end{bmatrix},$$

式中,A_1 为阶数不超过 n 的单位矩阵(或对应地,若 $\rho(A)=1$ 且条件(2)成立,则 A_1 为阶数不超 n 的对角矩阵,其对角元素的模均为 1),而 A_2 满足 $\rho(A_2)<1$. 于是,$\lim\limits_{k\to\infty}(P^{-1}AP)^k=\lim\limits_{k\to\infty}P^{-1}A^kP=\begin{bmatrix} I & O \\ O & O \end{bmatrix}$(若条件(2)成立,则$(P^{-1}AP)^k$ 当 $k\to\infty$ 时有界),此蕴涵$\lim\limits_{k\to\infty}A^k$ 存在(或对应地,A^k 有界). □

作为前面几个定理的应用,我们得到一些有用的推论.

推论 8 设 $A\in \boldsymbol{M}_n(\mathbb{C})$. 若对某种 $\boldsymbol{M}_n(\mathbb{C})$ 上矩阵范数 $\|\cdot\|$, $\|A\|<1$, 则$\lim\limits_{k\to\infty}A^k=O$.

证明 应用定理 4 与定理 6 即得. □

推论 9 设 $A\in \boldsymbol{M}_n(\mathbb{C})$. 则$\lim\limits_{k\to\infty}A^k=O$ 当且仅当

$$\lim_{k\to\infty}A^k\boldsymbol{x} = \boldsymbol{0}, \forall\, \boldsymbol{x}\in\mathbb{C}^n. \tag{10}$$

证明 设 $\|\cdot\|_\alpha$ 为$\mathbb{C}^n$ 上某向量范数,则 $\|\cdot\|_\alpha$ 导出的矩阵范数 $\|\cdot\|$ 满足

$$\|A^k\boldsymbol{x}\|_\alpha \leqslant \|A^k\| \ \|\boldsymbol{x}\|_\alpha, \qquad \forall\, \boldsymbol{x}\in\mathbb{C}^n,$$

因而 $A^k\to O$ 蕴涵(10)式. 反之,假定(10)式成立. 取 $\boldsymbol{x}=\boldsymbol{e}_j$($\mathbb{C}^n$ 的第 j 个单位坐标向量),则 $A^k\boldsymbol{e}_j=A^k$ 的第 j 列向量$\to\boldsymbol{0}(k\to\infty)$, $j=1,\cdots,n$,因而按定义,$A^k\to O(k\to\infty)$. □

有时我们需要当 $k\to\infty$时 A^k 元素$(A^k)_{ij}$ 的界,应用定理 6 可以得到一个有用的界.

推论 10 设 $A\in \boldsymbol{M}_n(\mathbb{C})$且 $\varepsilon>0$ 任意给定,则存在与 A,ε 有关的常数 $c=c(A,\varepsilon)$使得

$$|(A^k)_{ij}|\leqslant c[\rho(A)+\varepsilon]^k, \qquad k=1,2,\cdots;i,j=1,\cdots,n.$$

证明 令 $\widetilde{A}\equiv A/[\rho(A)+\varepsilon]$,则 $\rho(\widetilde{A})<1$,因而 $\widetilde{A}^k\to O(k\to\infty)$. 这时$\{\widetilde{A}^k\}$元素序列有界,因而有有限数 $c>0$ 满足$|(\widetilde{A}^k)_{ij}|\leqslant c$, $k=1,2,\cdots;i,j=1,\cdots,n$. □

除了(9′)式以外,$\boldsymbol{M}_n(\mathbb{C})$上矩阵范数与谱半径之间的另一个基本关系由下面定理给出.

定理 11 设 $\|\cdot\|$ 为任意的 $\boldsymbol{M}_n(\mathbb{C})$上矩阵范数,则

$$\rho(A)=\lim_{k\to\infty}\|A^k\|^{1/k}=\inf_{k\geqslant 1}\|A^k\|^{1/k}, \qquad \forall\, A\in\boldsymbol{M}_n(\mathbb{C}). \tag{11}$$

证明 由于$(\rho(A))^k=\rho(A^k)\leqslant\|A^k\|$,故我们有 $\rho(A)\leqslant\|A^k\|^{1/k}$, $k=1,2,\cdots$,因而 $\rho(A)\leqslant\inf\limits_{k\geqslant 1}\|A^k\|^{1/k}$. 另一方面,对任意给定 $\varepsilon>0$,矩阵 $\widetilde{A}\equiv A/[\rho(A)+\varepsilon]$有$\rho(\widetilde{A})<1$,因而 $\widetilde{A}^k\to O(k\to\infty)$. 于是,存在 $N=N(\varepsilon,A)$使得 $\|\widetilde{A}^k\|<1$, $\forall\, k\geqslant N$,此等价于 $\|A^k\|\leqslant[\rho(A)+\varepsilon]^k$, $\forall\, k\geqslant N$,因而

$$\|A^k\|^{1/k}\leqslant\rho(A)+\varepsilon, \qquad \forall\, k\geqslant N.$$

由此推出$\limsup_{k\to\infty}\|A^k\|^{1/k}\leqslant\rho(A)$. 但显然有$\inf_{k\geqslant 1}\|A^k\|^{1/k}\leqslant\limsup_{k\to\infty}\|A^k\|^{1/k}$，于是，

$$\rho(A)\leqslant\inf_{k\geqslant 1}\|A^k\|^{1/k}\leqslant\limsup_{k\to\infty}\|A^k\|^{1/k}\leqslant\rho(A).$$

因此，$\lim_{k\to\infty}\|A^k\|^{1/k}$存在且(11)式成立. □

现在讨论前面几个结果在矩阵的无穷级数$\sum_{k=1}^{\infty}A_k$收敛性问题中的应用.

设$A_k\in\boldsymbol{M}_n(\mathbb{C}),k=1,2,\cdots$. 若$\sum_{k=1}^{\infty}A_k$的部分和序列$\left\{\sum_{k=1}^{m}A_k\right\}_{m=1}^{\infty}$收敛于矩阵$C\in\boldsymbol{M}_n(\mathbb{C})$，则称矩阵级数$\sum_{k=1}^{\infty}A_k$**收敛**于矩阵$C$，记为

$$\sum_{k=1}^{\infty}A_k=C,$$

否则$\sum_{k=1}^{\infty}A_k$便是**发散**的. 类似地，这样的矩阵级数称为**绝对收敛**的，假如对某个(因而对任意)$\boldsymbol{M}_n(\mathbb{C})$上矩阵范数$\|\cdot\|$，数项级数$\sum_{k=1}^{\infty}\|A_k\|$收敛. 此时，$\sum_{k=1}^{\infty}A_k$收敛，并且有$\left\|\sum_{k=1}^{\infty}A_k\right\|\leqslant\sum_{k=1}^{\infty}\|A_k\|$.

对于矩阵级数，因矩阵之间的相乘的可能性，所以它有一些向量级数所未有的公式. 例如，我们有

$$B\left(\sum A_k\right)=\sum(BA_k),\qquad\left(\sum A_k\right)B=\sum(A_kB),$$

只要其中$\sum A_k$收敛且乘积有意义，这些可直接从AB作为A与B函数的连续性的事实推得. 此外，两个绝对收敛的矩阵级数可以逐项相乘，即有

$$\left(\sum B_l\right)\left(\sum A_k\right)=\sum B_lA_k,$$

假如有关乘积有意义. 这时，上式右端各项的次序为任意的. 这个事实的证明基本上与数项级数情形相仿.

这里只考虑与矩阵$A\in\boldsymbol{M}_n(\mathbb{C})$有关的级数$\sum_{k=0}^{\infty}A^k$的收敛性问题. 更一般的矩阵级数的收敛性问题将放在第三章 3.2.3 小节讨论.

定理 12(Neumann 引理)　设$A\in\boldsymbol{M}_n(\mathbb{C})$. 则$I_n-A$非奇异且矩阵级数$\sum_{k=0}^{\infty}A^k$收敛当且仅当$\rho(A)<1$. 此时，

$$(I_n-A)^{-1}=\sum_{k=0}^{\infty}A^k.\tag{12}$$

此外,如对某个满足 $\|I_n\|=1$ 的 $\boldsymbol{M}_n(\mathbb{C})$ 上矩阵范数 $\|\cdot\|$, $\|A\|<1$,则 $\sum_{k=0}^{\infty}A^k$ 绝对收敛,且有

$$\|I_n-A\|^{-1}\leqslant\|(I_n-A)^{-1}\|\leqslant\sum_{k=0}^{\infty}\|A^k\|\leqslant(1-\|A\|)^{-1}. \tag{13}$$

证明 令 $S_k=I_n+A+\cdots+A^k, k=0,1,\cdots$. 则

$$(I_n-A)S_k=I_n-A^{k+1}. \tag{14}$$

若 $\rho(A)<1$,则 I_n-A 的谱 $\sigma(I_n-A)$ 不含 0,因而 I_n-A 非奇异. 此时,(14)式蕴涵

$$S_k-(I_n-A)^{-1}=-(I_n-A)^{-1}A^{k+1}. \tag{15}$$

根据定理 6,当 $k\to\infty$ 时上式右端趋于零矩阵,因而矩阵级数 $\sum_{k=0}^{\infty}A^k$ 收敛且(12)式成立. 反之,若 $\sum_{k=0}^{\infty}A^k$ 收敛,按定义易得 $A^k\to O(k\to\infty)$,因而定理 6 断言 $\rho(A)<1$.
当 $\|A\|<1$ 时,$\rho(A)<1$,因而(12)式成立. 应用(3)式便得 $\|I_n-A\|^{-1}\leqslant\|(I_n-A)^{-1}\|=\left\|\sum_{k=0}^{\infty}A^k\right\|$. 再由 $\|A^k\|\leqslant\|A\|^k$ 看出, $\sum_{k=0}^{\infty}\|A^k\|$ 收敛,且以 $\sum_{k=0}^{\infty}\|A^k\|$ 为其强级数. 由于 $\|I_n\|=1$ 故后者之和为 $(1-\|A\|)^{-1}$,因而(13)式成立. □

我们指出,从(12)式与(15)式可以断言:若 $\|\cdot\|$ 为满足 $\|I_n\|=1$ 的 $\boldsymbol{M}_n(\mathbb{C})$ 上矩阵范数,且 $\|A\|<1$,则

$$\|(I_n-A)^{-1}-S_k\|=\left\|\sum_{j=k+1}^{\infty}A^j\right\|\leqslant\sum_{j=k+1}^{\infty}\|A\|^j=\frac{\|A\|^{k+1}}{1-\|A\|}. \tag{16}$$

取 $k=0$,上式给出 $\|(I_n-A)^{-1}-I_n\|\leqslant\|A\|(1-\|A\|)^{-1}$. 令 $R\equiv I_n-A$,这表明 $R\to I_n$ 蕴涵 $R^{-1}\to I_n$,换句话说,R^{-1} 作为 R 的函数在 $R=I_n$ 处连续. 一般地,我们有:A^{-1} 为 A 的连续函数,确切地说,若 $A\in\boldsymbol{M}_n(\mathbb{C})$ 非奇异,则对充分小的 $E\in\boldsymbol{M}_n(\mathbb{C})$, $A+E$ 非奇异,且 $\|(A+E)^{-1}-A^{-1}\|\to0(\|E\|\to0)$. 这个结果还表明: $\boldsymbol{M}_n(\mathbb{C})$ 中所有非奇异矩阵的集合为开集. 我们在下列基本结果中给出这个连续性的证明与有关的估计式.

定理 13(扰动引理) 设 $A\in\boldsymbol{M}_n(\mathbb{C})$ 非奇异. 若 $E\in\boldsymbol{M}_n(\mathbb{C})$ 满足条件 $\|E\|<1/\|A^{-1}\|$,这里, $\|\cdot\|$ 为满足 $\|I_n\|=1$ 的 $\boldsymbol{M}_n(\mathbb{C})$ 上某个矩阵范数,则 $A+E$ 非奇异,并且,

$$\|(A+E)^{-1}\|\leqslant\frac{\|A^{-1}\|}{1-\|A^{-1}\|\|E\|}, \tag{17}$$

$$\|(A+E)^{-1}-A^{-1}\|\leqslant\frac{\|A^{-1}\|^2\|E\|}{1-\|A^{-1}\|\|E\|}. \tag{18}$$

证明　按题设，$\| A^{-1}E \| \leqslant \| A^{-1} \| \| E \| < 1$，因而 $I_n + A^{-1}E$ 非奇异. 由于 $A+E=A(I_n+A^{-1}E)$，故 $A+E$ 作为非奇异矩阵的乘积也是非奇异的，且 $(A+E)^{-1}=(I_n+A^{-1}E)^{-1}A^{-1}$. 应用(13)式与(16)式($k=0$)到 $\| (I_n+A^{-1}E)^{-1} \|$ 与 $\| (I_n+A^{-1}E)^{-1}-I_n \|$ 的估计即得

$$\| (A+E)^{-1} \| \leqslant \frac{\| A^{-1} \|}{1-\| A^{-1}E \|} \leqslant \frac{\| A^{-1} \|}{1-\| A^{-1} \| \| E \|}$$

$$\| (A+E)^{-1}-A^{-1} \| = \| ((I_n+A^{-1}E)^{-1}-I_n)A^{-1} \| \leqslant \frac{\| A^{-1} \|^2 \| E \|}{1-\| A^{-1} \| \| E \|}. \quad \square$$

对满足 $\| I_n \| = 1$ 的 $\boldsymbol{M}_n(\mathbb{C})$ 上矩阵范数 $\|\cdot\|$，若定义

$$\boldsymbol{\kappa}(A) = \begin{cases} \| A^{-1} \| \ \| A \|, & \text{如果 } A \text{ 非奇异}, \\ \infty, & \text{如果 } A \text{ 奇异}, \end{cases} \tag{19}$$

并称之为(关于矩阵范数 $\|\cdot\|$ 与矩阵逆的)A 的**条件数**，则 $\boldsymbol{\kappa}(A) \geqslant 1$，$\forall A \in \boldsymbol{M}_n(\mathbb{C})$. 应用(18)式可给出计算矩阵逆时相对误差的一个上界：若 $\| E \| < 1/\| A^{-1} \|$，这里 $\|\cdot\|$ 满足 $\| I_n \| = 1$，则有

$$\frac{\| (A+E)^{-1}-A^{-1} \|}{\| A^{-1} \|} \leqslant \frac{\boldsymbol{\kappa}(A)}{1-\boldsymbol{\kappa}(A)(\| E \| / \| A \|)} \frac{\| E \|}{\| A \|}. \tag{20}$$

最后，我们顺便指出，若引入某算子范数 $\|\cdot\|$，则复结合代数 $\boldsymbol{M}_n(\mathbb{C})$ $(n \geqslant 1)$ (见第一章 1.1.1 小节)成为 n^2 维复 Banach 空间. 由于算子范数满足次可乘条件(1)，$\boldsymbol{M}_n(\mathbb{C})$ 有单位元 I_n 满足 $I_nA=AI_n=A$，$\forall A \in \boldsymbol{M}_n(\mathbb{C})$，且 $\| I_n \| = 1$，按复 Banach代数的定义，上述复 Banach 空间便是一个 n^2 维的复 Banach 代数. 现设$\mathscr{S} \neq \{\boldsymbol{0}\}$为复线性空间，$\dim \mathscr{S} = n < \infty$. $\mathscr{S}$上所有线性变换组成的复代数 $\mathscr{L}(\mathscr{S})$若赋以算子范数：

$$\| T \| = \sup_{\| x \|_\alpha = 1} \| Tx \|_\alpha, \forall T \in \mathscr{L}(\mathscr{S}) \tag{21}$$

式中 $\|\cdot\|_\alpha$ 为 $\mathscr{S}$上某向量范数，则 $\mathscr{L}(\mathscr{S})$也为 n^2 维复 Banach 代数. 当给定 $\mathscr{S}$的基 $\mathscr{B}$ 时，$\mathscr{L}(\mathscr{S})$同构于 $\boldsymbol{M}_n(\mathbb{C})$(见第一章 1.1.2 小节). 此外，复 Banach 代数 $\boldsymbol{M}_n(\mathbb{C})$中每个含有单位元 I_n 的子代数仍为复 Banach 代数，并且，可以证明：每一个有限维(维数$\leqslant n^2$)的复 Banach 代数必定同构于 $\boldsymbol{M}_n(\mathbb{C})$中某个子代数. 因此，$\boldsymbol{M}_n(\mathbb{C})$可视为一种最基本的有限维复 Banach 代数. 进一步内容可见文献[12]的第三部分.

习题 2.2.2

1. 试证：

(1) 实值函数 $\| A \| = n \max\limits_{i,j} |a_{ij}|$，$\forall A=(a_{ij}) \in \boldsymbol{M}_n(\mathbb{C})$定义 $\boldsymbol{M}_n(\mathbb{C})$上一个矩阵范数.

(2) (3)式中各不等式成立.

2. 试证：$\boldsymbol{M}_n(\mathbb{C})$上的广义矩阵范数

$$\| A \|_p = \Big(\sum_{i,j} | a_{ij} |^p \Big)^{1/p}, \qquad p \geqslant 1$$

当且仅当 $1\leqslant p\leqslant 2$ 时为 $\boldsymbol{M}_n(\mathbb{C})$ 上的矩阵范数.

3. 设 $\|\cdot\|$ 为 $\boldsymbol{M}_n(\mathbb{C})$ 上矩阵范数，且 $S\in\boldsymbol{M}_n(\mathbb{C})$ 非奇异. 已证明 $v(A)\equiv\|SAS^{-1}\|$, $\forall A\in\boldsymbol{M}_n(\mathbb{C})$ 为 $\boldsymbol{M}_n(\mathbb{C})$ 上矩阵范数(见定理 2). 试证：若 $\|\cdot\|$ 为 $\mathbb{C}^n$ 上向量范数 $\|\cdot\|_\alpha$ 导出的矩阵范数，则 $v(\cdot)$ 也是 $\mathbb{C}^n$ 上某向量范数导出的矩阵范数.

4. 设 $\|\cdot\|_F$ 为 $\boldsymbol{M}_n(\mathbb{C})$ 上 Frobenius 范数，而 $\|\cdot\|$ 为题 1(1) 中矩阵范数，试证：$\|A\|_F\leqslant\|A\|$, $\forall A\in\boldsymbol{M}_n(\mathbb{C})$.

5. 设 $p\geqslant 2$, $p^{-1}+q^{-1}=1$. 试证：由题 2 确定的广义矩阵范数 $\|\cdot\|_p$ 满足

$$\|AB\|_p\leqslant\min(\|A\|_p\|B\|_q,\|A\|_q\|B\|_p),\qquad \forall A,B\in\boldsymbol{M}_n(\mathbb{C}).$$

6. 设 $\|\cdot\|_1$, $\|\cdot\|_\infty$ 与 $\|\cdot\|_2$ 分别为 $\boldsymbol{M}_n(\mathbb{C})$ 上矩阵的列范数、行范数与谱范数. 试证：$\|A\|_2^2\leqslant\|A\|_\infty\|A\|_1$, $\forall A\in\boldsymbol{M}_n(\mathbb{C})$.

7. 设 $A\in\boldsymbol{M}_n(\mathbb{C})$ 有分块对角形式 $A=\operatorname{diag}[A_1,A_2,\cdots,A_k]$，这里，诸 A_j 为方阵. 试证：

$$\|A\|_2=\max_{1\leqslant j\leqslant k}\|A_j\|_2.$$

8. 设 $A\in\boldsymbol{M}_n(\mathbb{C})$，并且，

$$H=\begin{bmatrix} O & \mathrm{i}A \\ -\mathrm{i}A^* & O\end{bmatrix},\qquad \mathrm{i}^2=-1$$

为 $2n$ 阶 Hermite 矩阵. 试证：$\|H\|_2=\|A\|_2$.

9. 设 $A\in\boldsymbol{M}_n(\mathbb{C})$. 试证：当且仅当对应于 A 的满足 $|\lambda|=\rho(A)$ 的特征值 λ 的 Jordan 块都是一阶时，$\boldsymbol{M}_n(\mathbb{C})$ 上有矩阵范数 $\|\cdot\|$ 使得 $\rho(A)=\|A\|$.

10. 设 $\|\cdot\|_\beta$ 为 $\boldsymbol{M}_n(\mathbb{C})$ 上广义矩阵范数. 令

$$N(A)=\sup_{O\neq X\in\boldsymbol{M}_n(\mathbb{C})}\frac{\|AX\|_\beta}{\|X\|_\beta}.$$

试证：

(1) $N(\cdot)$ 为 $\boldsymbol{M}_n(\mathbb{C})$ 上的矩阵范数.

(2) 若 $\|\cdot\|_\beta$ 为 $\boldsymbol{M}_n(\mathbb{C})$ 上任一矩阵范数，则 $N(A)\leqslant\|A\|_\beta$, $\forall A\in\boldsymbol{M}_n(\mathbb{C})$.

(3) 当且仅当 $\|\cdot\|_\beta$ 为满足 $\|I_n\|_\beta=1$ 的 $\boldsymbol{M}_n(\mathbb{C})$ 上矩阵范数时，$N(A)=\|A\|_\beta$, $\forall A\in\boldsymbol{M}_n(\mathbb{C})$.

11. 试证：若 $v(\cdot)$ 为 $\boldsymbol{M}_n(\mathbb{C})$ 上准范数(参见习题 2.1.2 题 3 与题 4)，则 $\lim\limits_{k\to\infty}[v(A^k)]^{1/k}$ 存在，$\forall A\in\boldsymbol{M}_n(\mathbb{C})$，且

$$\lim_{k\to\infty}[v(A^k)]^{1/k}=\rho(A).$$

12. 试证：对 $n>1$，谱半径 $\rho(A)$ 不是 $\boldsymbol{M}_n(\mathbb{C})$ 上的矩阵范数与广义矩阵范数.

13. 设 $A\in\boldsymbol{M}_n(\mathbb{C})$ 非奇异，有特征值 $|\lambda_1|\geqslant|\lambda_2|\geqslant\cdots\geqslant|\lambda_n|$，且 $\|\cdot\|_\alpha$ 为任一 $\mathbb{C}^n$ 上的向量范数. 试证：

(1) $\max\limits_{\|x\|_\alpha=\|y\|_\alpha=1}\|A\boldsymbol{x}\|_\alpha/\|A\boldsymbol{y}\|_\alpha\geqslant|\lambda_1|/|\lambda_n|$.

(2) 对与 $\|\cdot\|_\alpha$ 相容的 $\boldsymbol{M}_n(\mathbb{C})$ 上矩阵范数 $\|\cdot\|$，有

$$\max_{\|x\|_\alpha=\|y\|_\alpha=1}\|A\boldsymbol{x}\|_\alpha/\|A\boldsymbol{y}\|_\alpha=\|A^{-1}\|\ \|A\|.$$

14. 设 $A\in\boldsymbol{M}_n(\mathbb{C})$ 非奇异，$A\boldsymbol{x}=\boldsymbol{b}(\boldsymbol{b}\neq\boldsymbol{0})$ 为线性方程组. 考虑扰动方程组 $(A+\Delta A)\boldsymbol{z}=\boldsymbol{b}+\Delta\boldsymbol{b}$. 假定 $\|\cdot\|_\alpha$ 为 $\mathbb{C}^n$ 上向量范数，$\|\cdot\|$ 为满足 $\|I_n\|=1$ 且与 $\|\cdot\|_\alpha$ 相容的 $\boldsymbol{M}_n(\mathbb{C})$ 上矩阵范数. 试

证：若 $\|\Delta A\|<1/\|A^{-1}\|$，则上述两个方程组解 $\boldsymbol{x}$ 与 $\boldsymbol{z}$ 满足

$$\frac{\|\boldsymbol{z}-\boldsymbol{x}\|_\alpha}{\|\boldsymbol{x}\|_\alpha}\leqslant\frac{\kappa(A)}{1-\kappa(A)\|\Delta A\|/\|A\|}\left[\frac{\|\Delta \boldsymbol{b}\|_\alpha}{\|\boldsymbol{b}\|_\alpha}+\frac{\|\Delta A\|}{\|A\|}\right],$$

式中，$\kappa(A)=\|A^{-1}\|\ \|A\|$ 为 A 的条件数.

2.3 关于向量范数与矩阵范数的进一步结果

作为前两节知识的补充与延伸，我们在本节研究关于向量范数与矩阵范数的进一步结果，其中包括：**对偶向量范数**的概念与性质、**绝对向量范数**的概念与性质、矩阵的**下界**、**最小矩阵范数**、**矩阵数值域**以及**矩阵数值半径**等.

2.3.1 对偶向量范数

设 $\boldsymbol{F}^n=\mathbb{C}^n$ 或 $\mathbb{R}^n$. 应用 $\boldsymbol{F}^n$ 上任意向量范数的单位闭球是以 $\boldsymbol{0}$ 为内点的均衡凸体的几何性质，我们可以引出与向量范数**对偶**的范数. 这种对偶的范数仍然为 F^n 上的向量范数，并与原先的向量范数有着密切的关系. 这样一来，我们还顺便地得到从旧的向量范数生成新的向量范数的又一种有用的方法.

定义 1 设 $\|\cdot\|$ 为 $\boldsymbol{F}^n$ 上的向量范数. 由下式确定的 $\boldsymbol{F}^n$ 上实值函数 $\|\cdot\|_D$：

$$\|\boldsymbol{y}\|_D\equiv\max_{0\neq x\in \boldsymbol{F}^n}\frac{|\boldsymbol{y}^*\boldsymbol{x}|}{\|\boldsymbol{x}\|}\equiv\max_{\|\boldsymbol{x}\|=1}|\boldsymbol{y}^*\boldsymbol{x}|,\ \forall\,\boldsymbol{y}\in\boldsymbol{F}^n \tag{1}$$

称为 $\|\cdot\|$ 的**对偶范数**(dual norm).

说明一下，对于固定 $\boldsymbol{y}\in\boldsymbol{F}^n$，$|\boldsymbol{y}^*\boldsymbol{x}|$ 为 $\boldsymbol{x}$ 的实值连续函数，因而它在紧集 $\{\boldsymbol{x}\in\boldsymbol{F}^n:\|\boldsymbol{x}\|=1\}$ 上达到最大值. 这表明(1)式中 $\|\cdot\|_D$ 确为 $\boldsymbol{F}^n$ 上的实值函数. 容易验证 $\|\cdot\|_D$ 满足向量范数的所有条件. 事实上，它满足非负性与齐次性条件是显然的. 若 $\boldsymbol{y}\neq\boldsymbol{0}$，则有 $\|\boldsymbol{y}\|_D>0$，这是因为

$$\|\boldsymbol{y}\|_D=\max_{\|\boldsymbol{x}\|=1}|\boldsymbol{y}^*\boldsymbol{x}|\geqslant\left|\boldsymbol{y}^*\frac{\boldsymbol{y}}{\|\boldsymbol{y}\|}\right|=\|\boldsymbol{y}\|_2^2/\|\boldsymbol{y}\|>0.$$

最后，即使实值函数 $\|\cdot\|$ 不满足三角不等式条件，对于 $\|\cdot\|_D$ 也有

$$\begin{aligned}\|\boldsymbol{y}+\boldsymbol{z}\|_D&=\max_{\|\boldsymbol{x}\|=1}|(\boldsymbol{y}+\boldsymbol{z})^*\boldsymbol{x}|\leqslant\max_{\|\boldsymbol{x}\|=1}|\boldsymbol{y}^*\boldsymbol{x}|+\max_{\|\boldsymbol{x}\|=1}|\boldsymbol{z}^*\boldsymbol{x}|\\&=\|\boldsymbol{y}\|_D+\|\boldsymbol{z}\|_D.\end{aligned}$$

因此，$\boldsymbol{F}^n$ 上的准范数(参见习题 2.1.2 题 3 与题 4) $\|\cdot\|$ 确定的实值函数 $\|\cdot\|_D$(见(1)式)也为 $\boldsymbol{F}^n$ 上的向量范数. 这说明定义 1 中 $\|\cdot\|$ 为"向量范数"的假定可换为"准范数"，即只需 $\|\cdot\|$ 有正性、齐次性与连续性.

若 $c\in\boldsymbol{F}$ 满足 $|c|=1$，则按 $\|\cdot\|$ 的齐次性，我们有

$$\begin{aligned}\max_{\|\boldsymbol{x}\|=1}|\boldsymbol{y}^*\boldsymbol{x}|&=\max_{\|\boldsymbol{x}\|=1}\max_{|c|=1}\mathrm{Re}c\boldsymbol{y}^*\boldsymbol{x}\\&=\max_{|c|=1}\max_{\|\boldsymbol{x}/c\|=1}\mathrm{Re}\boldsymbol{y}^*\boldsymbol{x}=\max_{\|\boldsymbol{x}\|=1}\mathrm{Re}\boldsymbol{y}^*\boldsymbol{x},\end{aligned}$$

于是，(1)式等价于

$$\|y\|_D=\max_{0\neq x\in F^n}\frac{\mathrm{Re}y^*x}{\|x\|}=\max_{\|x\|=1}\mathrm{Re}y^*x.$$

显然，上式或(1)式中“$\|x\|=1$”均可换为“$\|x\|\leqslant 1$”，等式仍成立.

从(1)式直接看出，若 $\|\cdot\|$ 为 F^n 上准范数，则

$$\begin{aligned}|y^*x|&\leqslant\|x\|\,\|y\|_D,\quad \forall x,y\in F^n,\\ |y^*x|&\leqslant\|x\|_D\,\|y\|,\quad \forall x,y\in F^n.\end{aligned}\tag{2}$$

这是 Cauchy-Schwarz 不等式的一种自然推广. 特殊地可以证实(见习题 2)，若 $\|\cdot\|_p(p\geqslant 1)$ 为 F^n 上 Hölder 范数(见 2.1.1 小节(4)式)，则它的对偶范数便为 $\|\cdot\|_q$，这里，$p^{-1}+q^{-1}=1$. 因此，这时(2)式便退化为 Hölder 不等式：

$$|y^*x|\leqslant\|x\|_p\,\|y\|_q,\quad \forall x,y\in F^n,$$

亦即

$$\left|\sum_{j=1}^n(x_j\bar{y}_j)\right|\leqslant\left(\sum_{j=1}^n|x_j|^p\right)^{1/p}\left(\sum_{j=1}^n|y_j|^q\right)^{1/q}.\tag{3}$$

下面例子考虑刚提到结论的极端情形.

例 2 讨论 F^n 上最大范数 $\|\cdot\|_\infty$ 与和范数 $\|\cdot\|_1$ 的对偶范数.(亦见习题 2.1.1 题 9)

首先有

$$|y^*x|\leqslant\sum_{j=1}^n|x_j\bar{y}_j|\leqslant\max_j|y_j|\sum_{j=1}^n|x_j|=\|y\|_\infty\|x\|_1,\forall x,y.\tag{4}$$

因此按定义 1，$(\|y\|_1)_D=\max\limits_{\|x\|_1=1}|y^*x|\leqslant\|y\|_\infty$ 与 $(\|y\|_\infty)_D=\max\limits_{\|x\|_\infty=1}|y^*x|\leqslant\|y\|_1$，$\forall y\in F^n$. 另一方面，对任意给定 y，若 $|y_k|=\|y\|_\infty$，则取 $x=(x_1,\cdots,x_n)^{\mathrm{T}}\in F^n$ 使得 $x_k=1$，其他 $x_j=0$，这时，$\|x\|_1=1$ 且 $|y^*x|=|y_k|=\|y\|_\infty$. 于是，$(\|\cdot\|_1)_D=\|\cdot\|_\infty$. 类似地，若给定 $y\neq 0$，取 $x=(x_1,\cdots,x_n)^{\mathrm{T}}\in F^n$ 使得 $x_j=y_j/|y_j|$(如 $y_j\neq 0$)，否则 $x_j=0$，这时，$\|x\|_\infty=1$ 且 $|y^*x|=\|y\|_1\,\|x\|_\infty=\|y\|_1$. 于是，$(\|\cdot\|_\infty)_D=\|\cdot\|_1$. 因此，$F^n$ 上的向量范数 $\|\cdot\|_\infty$ 与 $\|\cdot\|_1$ 是互为对偶的. □

从习题 2 看出，F^n 上的欧氏范数 $\|\cdot\|_2$ 为自对偶的范数：$(\|\cdot\|_2)_D=\|\cdot\|_2$. 下面结果进一步地指出，欧氏范数 $\|\cdot\|_2$ 是 F^n 上唯一的自对偶向量范数.

定理 3 设 $\|\cdot\|$ 为 F^n 上向量范数，$\|\cdot\|_D$ 为它的对偶范数，且 $c>0$ 为任意给定的正数. 则 $\|x\|=c\|x\|_D$，$\forall x\in F^n$ 当且仅当 $\|\cdot\|=\sqrt{c}\,\|\cdot\|_2$. 特殊地，$\|\cdot\|=\|\cdot\|_D$ 当且仅当 $\|\cdot\|=\|\cdot\|_2$.

证明 若 $\|\cdot\|=\sqrt{c}\,\|\cdot\|_2$，且 $x\in F^n$，则

$$\|x\|_D=\max_{\|y\|=1}|x^*y|=\max_{\|y\|_2=1}|x^*y/\sqrt{c}|=\frac{1}{\sqrt{c}}(\|x\|_2)_D=\frac{1}{\sqrt{c}}\|x\|_2=\frac{1}{c}\|x\|.$$

反之,设 $\|\cdot\|=c\|\cdot\|_D$,且 $\boldsymbol{x}\in\boldsymbol{F}^n$,则(2)式给出

$$\|\boldsymbol{x}\|_2^2=|\boldsymbol{x}^*\boldsymbol{x}|\leqslant\|\boldsymbol{x}\|\|\boldsymbol{x}\|_D=\|\boldsymbol{x}\|^2/c,$$

因而 $\|\boldsymbol{x}\|\geqslant\sqrt{c}\|\boldsymbol{x}\|_2$. 应用这个不等式可以建立相反的不等式. 事实上,对 $\boldsymbol{x}\neq\boldsymbol{0}$,

$$\|\boldsymbol{x}\|/c=\|\boldsymbol{x}\|_D=\max_{\boldsymbol{y}\neq\boldsymbol{0}}|\boldsymbol{x}^*\boldsymbol{y}/\|\boldsymbol{y}\||=\max_{\boldsymbol{y}\neq\boldsymbol{0}}(|\boldsymbol{x}^*\boldsymbol{y}/\|\boldsymbol{y}\|_2|\|\boldsymbol{y}\|_2/\|\boldsymbol{y}\|)$$

$$\leqslant\max_{\boldsymbol{y}\neq\boldsymbol{0}}(|\boldsymbol{x}^*\boldsymbol{y}/\|\boldsymbol{y}\|_2|/\sqrt{c}=(\|\boldsymbol{x}\|_2)_D/\sqrt{c}=\|\boldsymbol{x}\|_2/\sqrt{c},$$

式中用到 $\|\boldsymbol{y}\|_2/\|\boldsymbol{y}\|\leqslant1/\sqrt{c}$, $\forall\,\boldsymbol{y}\neq\boldsymbol{0}$ 与 $(\|\cdot\|_2)_D=\|\cdot\|_2$ 的事实. 因此, $\|\boldsymbol{x}\|\leqslant\sqrt{c}\|\boldsymbol{x}\|_2$, $\forall\,\boldsymbol{x}\in\boldsymbol{F}^n$. 联合两个相反不等式则得 $\|\boldsymbol{x}\|=\sqrt{c}\|\boldsymbol{x}\|_2$, $\forall\,\boldsymbol{x}\in\boldsymbol{F}^n$. 本定理的最后一个结论为前一结论取 $c=1$ 的特殊情形. □

在本小节最后部分,我们将要证明对偶范数的一个基本定理——**对偶性定理**(见定理 5). 它指出:若 $\|\cdot\|$ 为 $\boldsymbol{F}^n$ 上向量范数,则 $\|\cdot\|_{DD}\equiv(\|\cdot\|_D)_D=\|\cdot\|$,即向量范数的对偶范数的对偶等于原先的向量范数. 为证明这个定理,先引入一些预备知识.

在实或复的空间 $\boldsymbol{F}^n$ 中,含有集合 S 的所有闭凸集之交称为 S 的**闭凸包**(closed convex hull),记为$\overline{\mathrm{Co}}S$. 由于整个空间是闭凸的,故至少有一个包含 S 的闭凸集. 若 S 本身为闭凸的,则 $S=\overline{\mathrm{Co}}S$. 显然,也可用另一种方式定义$\overline{\mathrm{Co}}S$,即$\overline{\mathrm{Co}}S$ 为含 S 的所有下列半空间之交:

$$H(\boldsymbol{u},\nu)\equiv\{\boldsymbol{x}\in\boldsymbol{F}^n:\mathrm{Re}\boldsymbol{u}^*\boldsymbol{x}\leqslant\nu\},\tag{5}$$

式中,$\boldsymbol{u}\in\boldsymbol{F}^n$ 为某一向量,$\nu\in\mathbb{R}$. 此种半空间(5)由**实超平面**

$$P(\boldsymbol{u},\nu)\equiv\{\boldsymbol{x}\in\boldsymbol{F}^n:\mathrm{Re}\boldsymbol{u}^*\boldsymbol{x}=\nu\}.\tag{6}$$

界定,且含超平面(6).

若 $S\subset\boldsymbol{F}^n$ 为不空闭凸集,$\boldsymbol{x}\in\boldsymbol{F}^n$ 但 $\boldsymbol{x}\notin S$,则存在实超平面 $P(\boldsymbol{u},\nu)$(这里 $\boldsymbol{0}\neq\boldsymbol{u}\in\boldsymbol{F}^n$,$\nu\in\mathbb{R}$)**分离** S 与 $\boldsymbol{x}$,即有

$$\boldsymbol{y}\in S\Rightarrow\mathrm{Re}\boldsymbol{u}^*\boldsymbol{y}\geqslant\nu,$$

$$\mathrm{Re}\boldsymbol{u}^*\boldsymbol{x}<\nu.\tag{7}$$

事实上,取 $\|\cdot\|_2$ 为 $\boldsymbol{F}^n$ 上欧氏向量范数,并设 $\boldsymbol{x}_s\in S$ 使得 $\|\boldsymbol{x}-\boldsymbol{x}_s\|_2=\mathrm{dist}(\boldsymbol{x},S)>0$(见习题 2.1.2 题 7). 这时对任意 $\boldsymbol{y}\in S$ 与 $0<\lambda<1$,$\boldsymbol{x}_s+\lambda(\boldsymbol{y}-\boldsymbol{x}_s)\in S$. 因此,$0\leqslant\|\boldsymbol{x}-\boldsymbol{x}_s-\lambda(\boldsymbol{y}-\boldsymbol{x}_s)\|_2^2-\|\boldsymbol{x}-\boldsymbol{x}_s\|_2^2=2\lambda\mathrm{Re}(\boldsymbol{y}-\boldsymbol{x}_s)^*(\boldsymbol{x}_s-\boldsymbol{x})+\lambda^2\|\boldsymbol{y}-\boldsymbol{x}_s\|_2^2$. 这个不等式蕴涵 $\mathrm{Re}(\boldsymbol{y}-\boldsymbol{x}_s)^*(\boldsymbol{x}_s-\boldsymbol{x})\geqslant0$,否则我们可以取 $\lambda\in(0,1)$满足$0<\lambda<-2\mathrm{Re}(\boldsymbol{y}-\boldsymbol{x}_s)^*(\boldsymbol{x}_s-\boldsymbol{x})/\|\boldsymbol{y}-\boldsymbol{x}_s\|_2^2$ 使得 $\|\boldsymbol{x}-\boldsymbol{x}_s-\lambda(\boldsymbol{y}-\boldsymbol{x}_s)\|_2<\|\boldsymbol{x}-\boldsymbol{x}_s\|_2$,但这与前面结果抵触. 于是,

$$\mathrm{Re}\boldsymbol{y}^*(\boldsymbol{x}_s-\boldsymbol{x})\geqslant\mathrm{Re}\boldsymbol{x}_s^*(\boldsymbol{x}_s-\boldsymbol{x})>\mathrm{Re}\boldsymbol{x}^*(\boldsymbol{x}_s-\boldsymbol{x}),$$

这是因为 $\mathrm{Re}(\boldsymbol{x}_s-\boldsymbol{x})^*(\boldsymbol{x}_s-\boldsymbol{x})=\|\boldsymbol{x}_s-\boldsymbol{x}\|_2^2>0$. 此时,取 $\boldsymbol{u}=\boldsymbol{x}_s-\boldsymbol{x}$ 与 $\nu=\mathrm{Re}\boldsymbol{x}_s^*(\boldsymbol{x}_s-\boldsymbol{x})$,则有(7)式成立.

这样一来,我们证明了所谓**超平面分离定理**.

定理 4　设 $S\neq\varnothing$ 为 $\boldsymbol{F}^n$ 内的闭凸集，且 $\boldsymbol{x}\notin S$. 则存在一个实超平面 $P(\boldsymbol{u},\nu)$，$\boldsymbol{0}\neq\boldsymbol{u}\in\boldsymbol{F}^n$，$\nu\in\mathbb{R}$，分离 S 与 $\boldsymbol{x}$，即有(7)式成立.

特别地，当 $S\neq\varnothing$ 为 $\boldsymbol{F}^n$ 内闭凸锥时，上述定理中 $\nu\in\mathbb{R}$ 可取为 $\nu=0$（即实超平面 $P(\boldsymbol{u},0)$ 通过锥的顶点 $\boldsymbol{0}$). 这里称 $S\subset\boldsymbol{F}^n$ 为锥，假如对任意非负实数 α，$\alpha S\subset S$.

有了这些预备知识，我们可以证明下列的**对偶性定理**.

定理 5　设 $\mathcal{M}(\cdot)$ 为 $\boldsymbol{F}^n$ 上准范数，$\mathcal{M}^D$ 表示 $\mathcal{M}$ 的对偶范数，$\mathcal{M}^{DD}$ 表示 $\mathcal{M}^D$ 的**对偶范数**，且设

$$B\equiv\{\boldsymbol{x}\in\boldsymbol{F}^n:\mathcal{M}(\boldsymbol{x})\leqslant 1\},$$
$$B''\equiv\{\boldsymbol{x}\in\boldsymbol{F}^n:\mathcal{M}^{DD}(\boldsymbol{x})\leqslant 1\}.$$

则有 $\mathcal{M}^{DD}(\boldsymbol{x})\leqslant\mathcal{M}(\boldsymbol{x})$，$\forall\,\boldsymbol{x}\in\boldsymbol{F}^n$，与 $B\subset B''$. 特殊地，若 $\mathcal{M}$ 为 $\boldsymbol{F}^n$ 上向量范数，则 $\mathcal{M}^{DD}=\mathcal{M}$ 与 $B=B''$.

证明　对给定 $\boldsymbol{x}\in\boldsymbol{F}^n$，(2)式给出 $|\boldsymbol{y}^*\boldsymbol{x}|\leqslant\mathcal{M}(\boldsymbol{x})\mathcal{M}^D(\boldsymbol{y})$，$\forall\,\boldsymbol{y}\in\boldsymbol{F}^n$，因而

$$\mathcal{M}^{DD}(\boldsymbol{x})=\max_{\mathcal{M}^D(\boldsymbol{y})\leqslant 1}|\boldsymbol{y}^*\boldsymbol{x}|\leqslant\max_{\mathcal{M}^D(\boldsymbol{y})\leqslant 1}\mathcal{M}(\boldsymbol{x})\mathcal{M}^D(\boldsymbol{y})\leqslant\mathcal{M}(\boldsymbol{x}). \tag{8}$$

于是，$\mathcal{M}^{DD}(\boldsymbol{x})\leqslant\mathcal{M}(\boldsymbol{x})$，$\forall\,\boldsymbol{x}\in\boldsymbol{F}^n$. 由此则得 $B\subset B''$. 当 $\mathcal{M}$ 为 $\boldsymbol{F}^n$ 上的向量范数时，B 与 B'' 均为 $\boldsymbol{F}^n$ 内的不空闭凸集，从定理 4 我们推出：

若 $\boldsymbol{x}$ 属于所有包含 B 的半空间 $\{z\in\boldsymbol{F}^n:\mathrm{Re}z^*\,\boldsymbol{v}\leqslant 1\}$ $(\boldsymbol{0}\neq\boldsymbol{v}\in\boldsymbol{F}^n)$，则 $\boldsymbol{x}\in B$，即有 $\mathcal{M}(\boldsymbol{x})\leqslant 1$.（事实上，包含 B 的半空间 $\{z\in\boldsymbol{F}^n:\mathrm{Re}z^*\,\boldsymbol{v}\leqslant r\}$ $(\boldsymbol{0}\neq\boldsymbol{v}\in\boldsymbol{F}^n$ 与 $r\in R)$ (9) 中实数 r 必定非零，否则从 $B\subset\{z\in\boldsymbol{F}^n:\mathrm{Re}z^*\,\boldsymbol{v}\leqslant 0\}$ 的事实一定导出 $B\subset\{z\in\boldsymbol{F}^n:\mathrm{Re}z^*\,\boldsymbol{v}\geqslant 0\}$，因而 $B\subset\{z\in\boldsymbol{F}^n:\mathrm{Re}z^*\,\boldsymbol{v}=0\}$，但这与 $\boldsymbol{F}^n$ 内实超平面不能包含 B 的事实矛盾. 因此，(9)式中含 B 的半空间的表示是不失普遍性的）但从 $\mathcal{M}^D(\boldsymbol{y})=\max\limits_{x\in B}\mathrm{Re}\boldsymbol{y}^*\boldsymbol{x}$ 看出，$B\subset\{z\in\boldsymbol{F}^n:\mathrm{Re}z^*\,\boldsymbol{y}\leqslant 1\}$ 等价于 $\mathcal{M}^D(\boldsymbol{y})\leqslant 1$. 因此，命题(9)等价于

$$\{\mathcal{M}^D(\boldsymbol{y})\leqslant 1\Rightarrow\mathrm{Re}\boldsymbol{x}^*\,\boldsymbol{y}\leqslant 1\}\Rightarrow\boldsymbol{x}\in B,$$

这表明，$\mathcal{M}^{DD}(\boldsymbol{x})\leqslant 1$ 蕴涵 $\mathcal{M}(\boldsymbol{x})\leqslant 1$，因而 $\mathcal{M}(\boldsymbol{x})\leqslant\mathcal{M}^{DD}(\boldsymbol{x})$，$\forall\,\boldsymbol{x}\in\boldsymbol{F}^n$.　因此，

$$\mathcal{M}^{DD}(\boldsymbol{x})=\mathcal{M}(\boldsymbol{x}),\ \forall\,\boldsymbol{x}\in\boldsymbol{F}^n,\text{于是 } B=B''.$$ □

我们指出，本定理第二部分的证明也可由习题 6 的结果直接导出.

习题 2.3.1

1. 试证：若 $\alpha>0$ 与 $\|\cdot\|_\alpha\equiv\alpha\|\cdot\|$，这里 $\|\cdot\|$ 为 $\boldsymbol{F}^n$ 上向量范数，则 $(\|\cdot\|_\alpha)_D=\|\cdot\|_D/\alpha$.

2. 设 $p\geqslant 1$，$p^{-1}+q^{-1}=1$. 试证：$\boldsymbol{F}^n$ 上的 l_p 范数的对偶范数为 l_q 范数.

3. 设 $\|\cdot\|_\alpha$ 与 $\|\cdot\|_\beta$ 为 $\mathbb{C}^n$ 上两个给定的向量范数，且有正数 $c>0$ 使得 $\|\boldsymbol{x}\|_\alpha\leqslant c\|\boldsymbol{x}\|_\beta$，$\forall\,\boldsymbol{x}\in\mathbb{C}^n$. 试证：对它们的对偶范数有 $(\|\boldsymbol{x}\|_\beta)_D\leqslant c(\|\boldsymbol{x}\|_\alpha)_D$，$\forall\,\boldsymbol{x}\in\mathbb{C}^n$.

4. 设 $\|\cdot\|$ 为 $\mathbb{C}^n$ 上向量范数，$\boldsymbol{x}\in\mathbb{C}^n$ 给定. 我们称集合 $\{\boldsymbol{y}\in\mathbb{C}^n:\|\boldsymbol{y}\|_D\|\boldsymbol{x}\|=\boldsymbol{y}^*\boldsymbol{x}=1\}$ 为 $\boldsymbol{x}$ **关于 $\|\cdot\|$ 的对偶**. 一个有序向量对 $(\boldsymbol{x},\boldsymbol{y})\in\mathbb{C}^n\times\mathbb{C}^n$ 称为关于 $\|\cdot\|$ 的**对偶对**，假如 $\boldsymbol{y}$ 属于 $\boldsymbol{x}$ 关于 $\|\cdot\|$ 的对偶. 试证：

(1) 零向量关于 $\|\cdot\|$ 的对偶是无定义的.

(2) 若 $\|\cdot\|$ 为$\mathbb{C}^n$ 上欧氏范数 $\|\cdot\|_2$,且 $\mathbf{0}\neq \boldsymbol{x}\in\mathbb{C}^n$ 给定,则 $\boldsymbol{x}$ 关于 $\|\cdot\|_2$ 的对偶恰为$\{\boldsymbol{x}\}$.

(3) 若 $\|\cdot\|$ 为$\mathbb{C}^2$ 上范数 $\|\cdot\|_\infty$,则 $\boldsymbol{x}=(0,1)^{\mathrm{T}}$ 关于 $\|\cdot\|_\infty$的对偶恰为$\{\boldsymbol{x}\}$,但 $\boldsymbol{x}=(1,1)^{\mathrm{T}}$ 关于 $\|\cdot\|_\infty$的对偶为集合$\{(\alpha,1-\alpha)^{\mathrm{T}}:0\leqslant\alpha\leqslant1\}$.

(4) 若 $\|\cdot\|$ 为$\mathbb{C}^n$ 上向量范数,且 $\mathbf{0}\neq \boldsymbol{x}\in\mathbb{C}^n$ 给定,则 $\boldsymbol{y}$ 属于 $\boldsymbol{x}$ 关于 $\|\cdot\|$ 的对偶蕴涵 $\boldsymbol{x}$ 必属于 $\boldsymbol{y}$ 关于$(\|\cdot\|)_D$ 的对偶.

(5) $\mathbf{0}\neq\boldsymbol{x}$ 关于 $\|\cdot\|$ 的对偶等于$\{\boldsymbol{x}\}$当且仅当 $\|\cdot\|=\|\cdot\|_2$.

5. 设 $\boldsymbol{y}\in\mathbb{C}^n$, $\|\cdot\|$ 为$\mathbb{C}^n$ 上向量范数. 试证:存在 $\boldsymbol{y}_0\in\mathbb{C}^n$ 使得 $|\boldsymbol{y}_0^*\boldsymbol{x}|\leqslant\|\boldsymbol{x}\|$ $(\forall\boldsymbol{x}\in\mathbb{C}^n)$与 $\boldsymbol{y}_0^*\boldsymbol{y}=\|\boldsymbol{y}\|$.(此命题表明,任意 $\boldsymbol{x}(\neq\mathbf{0})$关于 $\|\cdot\|$ 的对偶是不空的)

6. 试证:若 $\mathscr{M}(\cdot)$为 $\boldsymbol{F}^n$ 上准范数,则 $B\subset B''\subset\overline{\mathrm{Co}}B$,这里 B 与 B''涵义同定理 5.

2.3.2　绝对向量范数及其导出的矩阵范数

前面讨论的向量范数中不少只与向量元素的模有关,例如,$l_p(p\geqslant1)$向量范数有性质:$\|\boldsymbol{x}\|_p=\||\boldsymbol{x}|\|_p$, $\forall\boldsymbol{x}\in\boldsymbol{F}^n$,这里,若 $\boldsymbol{x}=(x_1,\cdots,x_n)^{\mathrm{T}}\in\boldsymbol{F}^n$,则 $|\boldsymbol{x}|=(|x_1|,\cdots,|x_n|)^{\mathrm{T}}\in\mathbb{R}^n$. 而且不少向量范数为 $\boldsymbol{x}$ 元素模的单调增函数,例如,$l_p(p\geqslant1)$向量范数满足:只要 $|\boldsymbol{x}|\leqslant|\boldsymbol{y}|$(即 $|x_1|\leqslant|y_1|,\cdots,|x_n|\leqslant|y_n|$)便有 $\|\boldsymbol{x}\|_p\leqslant\|\boldsymbol{y}\|_p$,这里 $\boldsymbol{x},\boldsymbol{y}\in\boldsymbol{F}^n$. 但是,由矩阵谱范数 $\|A\|_2=(\rho(A^*A))^{1/2}$定义的 $\boldsymbol{F}^{n^2}$ 上向量范数不具有前述的两个性质.

定义 1　$\boldsymbol{F}^n$ 上向量范数 $\|\cdot\|$ 叫作**绝对的**(absolute),假如 $\|\boldsymbol{x}\|=\||\boldsymbol{x}|\|$, $\forall\boldsymbol{x}\in\boldsymbol{F}^n$;向量范数 $\|\cdot\|$ 叫作**单调的**(monotone),假如 $|\boldsymbol{x}|\leqslant|\boldsymbol{y}|$ 蕴涵 $\|\boldsymbol{x}\|\leqslant\|\boldsymbol{y}\|$,这里 $\boldsymbol{x},\boldsymbol{y}\in\boldsymbol{F}^n$.

下面基本定理表明,$\boldsymbol{F}^n$ 上向量范数的绝对性等价于它的对偶范数的绝对性.

定理 2　$\boldsymbol{F}^n$ 上任一绝对向量范数 $\|\cdot\|$ 的对偶范数 $\|\cdot\|_D$ 也是绝对的.

证明　对任意给定的 $\boldsymbol{y}$ 与 $\boldsymbol{x}$,存在 $\tilde{\boldsymbol{x}}\in\boldsymbol{F}^n$ 使得 $|\tilde{\boldsymbol{x}}|=|\boldsymbol{x}|$ 与 $\boldsymbol{y}^*\tilde{\boldsymbol{x}}=|\boldsymbol{y}|^{\mathrm{T}}|\boldsymbol{x}|$. 此时有

$$\|\boldsymbol{y}\|_D=\max_{x\neq\mathbf{0}}|\boldsymbol{y}^*\boldsymbol{x}|/\|\boldsymbol{x}\|\geqslant\max_{x\neq\mathbf{0}}\boldsymbol{y}^*\tilde{\boldsymbol{x}}/\|\tilde{\boldsymbol{x}}\|.$$

用 $|\boldsymbol{y}|^{\mathrm{T}}|\boldsymbol{x}|$代替上式中 $\boldsymbol{y}^*\tilde{\boldsymbol{x}}$,用 $\|\boldsymbol{x}\|$ 代替 $\|\tilde{\boldsymbol{x}}\|$(因为 $\|\cdot\|$ 为绝对的),则得出

$$\|\boldsymbol{y}\|_D\geqslant\max_{x\neq\mathbf{0}}|\boldsymbol{y}|^{\mathrm{T}}|\boldsymbol{x}|/\|\boldsymbol{x}\|.$$

另一方面,

$$\max_{x\neq\mathbf{0}}|\boldsymbol{y}^*\boldsymbol{x}|/\|\boldsymbol{x}\|\leqslant\max_{x\neq\mathbf{0}}|\boldsymbol{y}^*||\boldsymbol{x}|/\|\boldsymbol{x}\|=\max_{x\neq\mathbf{0}}|\boldsymbol{y}|^{\mathrm{T}}|\boldsymbol{x}|/\|\boldsymbol{x}\|.$$

因此,

$$\|\boldsymbol{y}\|_D=\max_{x\neq\mathbf{0}}|\boldsymbol{y}^*||\boldsymbol{x}|/\|\boldsymbol{x}\|. \tag{1}$$

这表明 $\|\boldsymbol{y}\|_D$ 只依赖于 $|\boldsymbol{y}|$,即 $\|\cdot\|_D$ 为绝对的. □

我们指出,(1)式蕴涵:若 $\|\cdot\|$ 为 $\boldsymbol{F}^n$ 上绝对向量范数,则有

$$|\boldsymbol{y}^*||\boldsymbol{x}|\leqslant\|\boldsymbol{x}\|\ \|\boldsymbol{y}\|_D,\qquad\forall\boldsymbol{x},\boldsymbol{y}\in\boldsymbol{F}^n. \tag{2}$$

此式通常叫作**广义的强 Cauchy-Schwarz 不等式**.

由对偶性定理(2.3.1 小节定理 5)知道,当 $\|\cdot\|$ 为 $\boldsymbol{F}^n$ 上绝对向量范数时,下列关系式也成立:

$$\|\boldsymbol{x}\| = \max_{y\neq 0} |\boldsymbol{y}^*| |\boldsymbol{x}| / \|\boldsymbol{y}\|_D, \forall \boldsymbol{x} \in \boldsymbol{F}^n. \tag{3}$$

应用定理 2,我们可以得到向量范数绝对性的基本特征,即绝对性等价于单调性.

定理 3 $\boldsymbol{F}^n$ 上向量范数 $\|\cdot\|$ 是绝对的当且仅当它是单调的.

证明 设 $\|\cdot\|$ 是绝对的. 按定理 2, $\|\cdot\|$ 的对偶范数 $\|\cdot\|_D$ 也是绝对的. 从(3)式知道,存在 $\tilde{\boldsymbol{u}}\neq\boldsymbol{0}$ 使得

$$\|\boldsymbol{x}\| = \max_{u\neq 0} |\boldsymbol{u}^*| |\boldsymbol{x}| / \|\boldsymbol{u}\|_D = |\tilde{\boldsymbol{u}}^*| |\boldsymbol{x}| / \|\tilde{\boldsymbol{u}}\|_D.$$

现若 $|\boldsymbol{x}|\leqslant|\boldsymbol{y}|$,则从上式有

$$\|\boldsymbol{x}\| \leqslant |\tilde{\boldsymbol{u}}^*| |\boldsymbol{y}| / \|\tilde{\boldsymbol{u}}\|_D \leqslant \max_{u\neq 0} |\boldsymbol{u}^*| |\boldsymbol{y}| / \|\boldsymbol{u}\|_D = \|\boldsymbol{y}\|.$$

因此, $\|\cdot\|$ 为单调的. 反之,若 $\|\cdot\|$ 单调,令 $\boldsymbol{y}=|\boldsymbol{x}|$,则有 $|\boldsymbol{x}|=|\boldsymbol{y}|$,因而 $\|\boldsymbol{x}\|\leqslant\|\boldsymbol{y}\|=\||\boldsymbol{x}|\|\leqslant\|\boldsymbol{x}\|$. 因此, $\|\boldsymbol{x}\|=\||\boldsymbol{x}|\|, \forall\boldsymbol{x}\in\boldsymbol{F}^n$,即 $\|\cdot\|$ 为绝对的. □

我们顺便地指出,可以不必应用定理 2 而从绝对向量范数的定义直接证明定理 3 中条件的必要性. 事实上,当 $\|\cdot\|$ 绝对时,为证 $\|\cdot\|$ 的单调性只需验证:若 $\boldsymbol{x},\boldsymbol{y}\in\boldsymbol{F}^n$ 满足 $0\leqslant x_j\leqslant y_j\,(j=1,\cdots,n)$,则有 $\|\boldsymbol{x}\|\leqslant\|\boldsymbol{y}\|$. 并且,为证明这个命题,我们实际上只需在 $0\leqslant x_1<y_1=1, 0\leqslant x_2=y_2,\cdots,0\leqslant x_n=y_n$ 假定下推出 $\|\boldsymbol{x}\|\leqslant\|\boldsymbol{y}\|$. 为此,令 $f(\boldsymbol{x})=f(x_1,\cdots,x_n)=\|\boldsymbol{x}\|, \forall\boldsymbol{x}\in\boldsymbol{F}^n$,并考虑以下分解式

$$(x_1,y_2,\cdots,y_n)=\frac{1+x_1}{2}(1,y_2,\cdots,y_n)+\frac{1-x_1}{2}(-1,y_2,\cdots,y_n).$$

应用向量范数公理条件与 $\|\cdot\|$ 的绝对性,我们有

$$\begin{aligned}\|\boldsymbol{x}\| &= f(x_1,\cdots,x_n)\leqslant\frac{1+x_1}{2}f(1,y_2,\cdots,y_n)+\frac{1-x_1}{2}f(-1,y_2,\cdots,y_n)\\ &=\frac{1+x_1}{2}f(1,y_2,\cdots,y_n)+\frac{1-x_1}{2}f(1,y_2,\cdots,y_n)\\ &=f(1,y_2,\cdots,y_n)=\|\boldsymbol{y}\|.\end{aligned}$$

□

$\mathbb{C}^n$ 上绝对向量范数的另一个特性是关于它导出的矩阵范数的性质. 由前面知识知道, $\mathbb{C}^n$ 上(绝对)向量范数 $\|\cdot\|_1$, $\|\cdot\|_2$ 与 $\|\cdot\|_\infty$ 导出的矩阵范数,即列范数,谱范数与行范数,在 n 阶复对角矩阵 $D=\mathrm{diag}[d_1,d_2,\cdots,d_n]$ 处的值均为 $\max\limits_j|d_j|$. 一般地有

定理 4 设 $\|\cdot\|$ 为 $\mathbb{C}^n$ 上的向量范数. 则 $\|\cdot\|$ 是绝对的当且仅当 $\|\cdot\|$ 导出的矩阵范数 $\|\cdot\|_\beta$ 对 $\boldsymbol{M}_n(\mathbb{C})$ 中所有对角矩阵 $D=\mathrm{diag}[d_1,d_2,\cdots,d_n]$ 有 $\|D\|_\beta=\max\limits_j|d_j|$.

证明 显然, $|D\boldsymbol{x}|\leqslant\max\limits_j|d_i|\,|\boldsymbol{x}|, \forall\boldsymbol{x}\in\mathbb{C}^n$ 与 $\forall D=\mathrm{diag}[d_1,d_2,\cdots,d_n]$. 现

设 $\|\cdot\|$ 为绝对的，则按定理 3，$\|Dx\|\leqslant\max_j|d_j|\ \|x\|$，$\forall x\in\mathbb{C}^n$，因而 $\|D\|_\beta\leqslant\max_j|d_j|$. 另一方面，$De_j=d_je_j$，$j=1,2,\cdots,n$，这里 e_j 为$\mathbb{C}^n$ 的第 j 个单位坐标向量. 按导出的矩阵范数定义，$|d_j|\ \|e_j\|=\|De_j\|\leqslant\|D\|_\beta\|e_j\|$，因而 $|d_j|\leqslant\|D\|_\beta$，$j=1,2,\cdots,n$. 于是 $\|D\|_\beta=\max_j|d_j|$. 反之，设 $\|D\|_\beta=\max_j|d_j|$ 对所有 $D=\mathrm{diag}[d_1,d_2,\cdots,d_n]\in\boldsymbol{M}_n(\mathbb{C})$ 成立. 对任意给定 $x\in\mathbb{C}^n$，有对角矩阵 $D_x\in\boldsymbol{M}_n(\mathbb{C})$ 使得 $x=D_x|x|$ 与 $|D_x|=I_n$，这里 $|D_x|$ 表示以 D_x 元素模为元素的 n 阶实矩阵. 此时，从 $\|D_x\|_\beta=\|D_x^{-1}\|_\beta=1$ 推出，

$$\|x\|=\|D_x|x|\|\leqslant\|D_x\|_\beta\||x|\|=\||x|\|,$$
$$\||x|\|=\|D_x^{-1}x\|\leqslant\|D_x^{-1}\|_\beta\|x\|=\|x\|.$$

因此，$\|x\|=\||x|\|$，$\forall x\in\mathbb{C}^n$. □

我们指出，$\mathbb{C}^n$ 上单调(绝对)向量范数 $\|\cdot\|$ 导出的矩阵范数 $\|\cdot\|_\beta$ 不必是单调的，换句话说，若 $A=(a_{ij})$，$B=(b_{ij})\in\boldsymbol{M}_n(\mathbb{C})$ 满足 $|A|\leqslant|B|$（即 $|a_{ij}|\leqslant|b_{ij}|$，$\forall i,j$)，则不必有 $\|A\|_\beta\leqslant\|B\|_\beta$. 例如，对 $\boldsymbol{M}_n(\mathbb{C})$ 上谱范数 $\|\cdot\|_2$，它显然不是单调的. 但是，当 $\|\cdot\|$ 为 $\mathbb{C}^n$ 上单调向量范数时它导出的矩阵范数 $\|\cdot\|_\beta$ 的如下一种弱单调性仍成立(见习题 3)：

$O\leqslant A\leqslant B\Rightarrow\|A\|_\beta\leqslant\|B\|_\beta$. 并且，$\|A\|_\beta\leqslant\||A|\|_\beta$，$\forall A\in\boldsymbol{M}_n(\mathbb{C})$，这里，$|A|=(|a_{ij}|)$，$A\leqslant B$ 意味着 $a_{ij}\leqslant b_{ij}$，$\forall i,j$.

最后，我们考虑**矩阵下界**的概念与有关性质.

从前面知道，$\mathbb{C}^n$ 上向量范数 $\|\cdot\|$ 导出的矩阵范数 $\|\cdot\|_\beta$ 由 $\|A\|_\beta=\sup_{\|x\|=1}\|Ax\|$ 定义. 特别地，当 S 为$\mathbb{C}^n$ 内有内点 $\mathbf{0}$ 的均衡凸体时，它对应$\mathbb{C}^n$ 上一个向量范数

$$\|x\|_S=\inf\{\nu:\nu\geqslant 0,x\in\nu S\}\tag{4}$$

(它为 2.1.2 小节(6)式的等价表达式)，$\|\cdot\|_S$ 导出的矩阵范数 $\sup_{\|x\|_S=1}\|Ax\|_S$ 有明显表示：$\mathrm{lub}_S(A)\equiv\inf\{\alpha:\alpha\geqslant 0,AS\subset\alpha S\}$. 事实上，容易验证 $\mathrm{lub}_S(\cdot)$ 满足 $\boldsymbol{M}_n(\mathbb{C})$ 上导出的矩阵范数所有条件：首先，$\|Ax\|_S\leqslant\mathrm{lub}_S(A)\|x\|_S$，$\forall x\in\mathbb{C}^n$. 其次，对任何 $A\in\boldsymbol{M}_n(\mathbb{C})$ 至少有一个 $x_0\neq\mathbf{0}$ 使得 $\|Ax_0\|_S=\mathrm{lub}_S(A)\|x_0\|_S$. (见习题 4)因此，通常将 $\mathrm{lub}_S(A)$ 称为 A 对于 S 的**最小上界**(简称为**界**).

同矩阵的(最小上)界相关的是与以 $\mathbf{0}$ 为内点的均衡凸体 S 对应的**最大下界**(简称**矩阵下界**)：

$$\mathrm{glb}_S(A)\equiv\sup\{\alpha:\alpha\geqslant 0,\alpha S\subset AS\},\qquad\forall A\in\boldsymbol{M}_n(\mathbb{C}).\tag{5}$$

可以验证(见习题 5)，

$$\|Ax\|_S\geqslant\mathrm{glb}_S(A)\|x\|_S,$$

且有 $x_0\neq\mathbf{0}$ 使得 $\|Ax_0\|_S=\mathrm{glb}_S(A)\|x_0\|_S$. 因此，

$$\mathrm{glb}_S(A)=\inf_{0\neq x\in\mathbb{C}^n}\|Ax\|_S/\|x\|_S. \tag{6}$$

由于$\mathbb{C}^n$上向量范数$\|\cdot\|_S$与均衡凸体S之间一一对应，故(6)式中$\|\cdot\|_S$可视为$\mathbb{C}^n$上任意给定的向量范数. 今后，若无混淆情况，常将(6)式中下标“S”去掉.

矩阵下界可用作一个矩阵关于给定向量范数的**非奇异性**的度量(更确切地说，与奇异性偏离的度量)，因为对奇异矩阵$A\in\boldsymbol{M}_n(\mathbb{C})$，$\mathrm{glb}_S(A)=0$，反之亦然；对接近奇异矩阵的非奇异矩阵$A\in\boldsymbol{M}_n(\mathbb{C})$，如承认这时$\mathrm{lub}_S(A^{-1})$很大，则其倒数$\mathrm{glb}_S(A)$很小. 确切地有如下结果.

定理 5　设$A\in\boldsymbol{M}_n(\mathbb{C})$，且$\|\cdot\|_\beta$为$\mathbb{C}^n$上向量范数$\|\cdot\|$导出的矩阵范数. 则有

$$\mathrm{glb}(A)=\begin{cases}0, & \text{如果 } \det A=0,\\ 1/\|A^{-1}\|_\beta, & \text{如果 } \det A\neq 0.\end{cases} \tag{7}$$

证明　若$\det A=0$，则有$\boldsymbol{0}\neq\boldsymbol{y}\in\mathbb{C}^n$使得$A\boldsymbol{y}=\boldsymbol{0}$，因此(6)式给出$\mathrm{glb}(A)=0$. 若$\det A\neq 0$，令$\boldsymbol{y}=A\boldsymbol{x}$，则对任意$\boldsymbol{x}\neq\boldsymbol{0}$有$\boldsymbol{y}\neq\boldsymbol{0}$，

$$\mathrm{glb}(A)=\inf_{x\neq 0}\frac{\|A\boldsymbol{x}\|}{\|\boldsymbol{x}\|}=\inf_{y\neq 0}\frac{\|\boldsymbol{y}\|}{\|A^{-1}\boldsymbol{y}\|}=\left[\sup_{y\neq 0}\frac{\|A^{-1}\boldsymbol{y}\|}{\|\boldsymbol{y}\|}\right]^{-1}=\frac{1}{\|A^{-1}\|_\beta}.$$

因此(7)式成立. □

定理 5 指出，$A\in\boldsymbol{M}_n(\mathbb{C})$非奇异当且仅当对某向量范数，$\mathrm{glb}(A)>0$. 此外，由(7)式容易推出

$$\mathrm{glb}(AB)\geqslant\mathrm{glb}(A)\mathrm{glb}(B),\qquad\forall A,B\in\boldsymbol{M}_n(\mathbb{C}). \tag{8}$$

从实用观点来看，由(6)式定义的矩阵非奇异度量不是十分令人满意的，因为它一般要计算矩阵的逆及其范数. 但是，在某些情形，它们是容易计算的.

推论 6　若$\|\cdot\|$为$\mathbb{C}^n$上绝对向量范数，且$D=\mathrm{diag}[d_1,d_2,\cdots,d_n]\in\boldsymbol{M}_n(\mathbb{C})$，则

$$\mathrm{glb}(D)=\min_j|d_j|. \tag{9}$$

证明　除去$\det D=0$平凡情形，考虑D非奇异情形，这时所有$d_j\neq 0$. 应用定理 5 与定理 4，

$$\mathrm{glb}(D)=\frac{1}{\|D^{-1}\|_\beta}=\frac{1}{\max_j(1/|d_j|)}=\min_j|d_j|.$$ □

推论 7　若$A\in\boldsymbol{M}_n(\mathbb{C})$正规，则与$\mathbb{C}^n$上欧氏范数$\|\cdot\|_2$对应的矩阵下界满足

$$\mathrm{glb}(A)=\min\{|\lambda|:\lambda\in\sigma(A)\}.$$

证明　当A奇异时，由(7)式知本推论成立. 现设$\lambda_1,\lambda_2,\cdots,\lambda_n$为非奇异正规矩阵$A$的特征值，则有酉矩阵$U$使得$A=U\mathrm{diag}[\lambda_1,\lambda_2,\cdots,\lambda_n]U^*$. 应用谱范数的酉不变性，$\|A^{-1}\|_2=\|\mathrm{diag}[\lambda_1^{-1},\lambda_2^{-1},\cdots,\lambda_n^{-1}]\|_2=\max_j|\lambda_j^{-1}|$. 再应用(7)式便得所要的结论. □

但是，对一般矩阵$A\in\boldsymbol{M}_n(\mathbb{C})$，只有如下结果.

推论 8 若 $A\in \boldsymbol{M}_n(\mathbb{C})$，则对应于$\mathbb{C}^n$ 上任一向量范数 $\|\cdot\|$，

$$\text{glb}(A)\leqslant \min\{|\lambda|:\lambda\in\sigma(A)\}.$$

证明 对 $\lambda\in\sigma(A)$ 有 $\boldsymbol{u}\neq\boldsymbol{0}$ 满足 $A\boldsymbol{u}=\lambda\boldsymbol{u}$. 此时有 $\|A\boldsymbol{u}\|\geqslant \text{glb}(A)\|\boldsymbol{u}\|$，但 $\|A\boldsymbol{u}\|=|\lambda|\|\boldsymbol{u}\|$. □

我们顺便地指出，应用推论 8 与 2.2.2 小节定理 4 可以确定出复平面上一个圆环区域，它的内部与边界包含 $\sigma(A)$.

习题 2.3.2

1. 试证：若 $A=\boldsymbol{a}\boldsymbol{b}^*$，$\boldsymbol{a},\boldsymbol{b}\in\mathbb{C}^n$，且 $\|\cdot\|_\beta$ 与 $\|\cdot\|_D$ 分别表示$\mathbb{C}^n$ 上向量范数 $\|\cdot\|$ 导出的矩阵范数与对偶范数，则有 $\|A\|_\beta=\|\boldsymbol{a}\|\|\boldsymbol{b}\|_D$.

2. 设$\mathbb{C}^n$ 上向量范数 $\|\cdot\|$ 的对偶范数为 $\|\cdot\|_D$. 试证：$\|\cdot\|$ 为绝对的当且仅当对所有 $\boldsymbol{x},\boldsymbol{y}\in\mathbb{C}^n$，广义的强 Cauchy-Schwarz 不等式(2)成立.

3. 设 $\|\cdot\|$ 为$\mathbb{C}^n$ 上绝对向量范数，其导出矩阵范数为 $\|\cdot\|_\beta$. 试证：$\|\cdot\|_\beta$ 满足正卦限中的单调性，即

$$O\leqslant A\leqslant B\Rightarrow \|A\|_\beta\leqslant\|B\|_\beta.$$

(这里，$A\leqslant B$ 意味着 $a_{ij}\leqslant b_{ij}$，$\forall\, i,j$)并且，由 $\nu(A)\equiv|||A|||_\beta$ 定义的实值函数为 $\boldsymbol{M}_n(\mathbb{C})$上的矩阵范数，它满足 $\|A\|_\beta\leqslant|||A|||_\beta$，$\forall A\in\boldsymbol{M}_n(\mathbb{C})$，式中，$|A|=(|a_{ij}|)$.

4. 设 $\|\cdot\|_S$ 由(4)式定义. 试验证：

$$\sup_{\|x\|_S=1}\|A\boldsymbol{x}\|_S=\text{lub}_S(A)\equiv\inf\{\alpha\geqslant 0:AS\subset\alpha S\}.$$

5. 试验证(6)式.

6. 设 $\|\cdot\|$ 为 $\boldsymbol{F}^n$ 上向量范数. 试证：$\|\cdot\|$ 为绝对的当且仅当 $\|(x_1,x_2,\cdots,x_n)^{\mathrm{T}}\|=\|(\alpha_1x_1,\alpha_2x_2,\cdots,\alpha_nx_n)^{\mathrm{T}}\|$ 对所有$(x_1,x_2,\cdots,x_n)^{\mathrm{T}}\in\boldsymbol{F}^n$ 与模为 1 的 $\alpha_1,\alpha_2,\cdots,\alpha_n\in\boldsymbol{F}$ 成立.

7. 设 $A\in\boldsymbol{M}_n(\mathbb{C})$非奇异. 试证：存在$\mathbb{C}^n$ 上向量范数使得对应的 $\text{glb}(A)=\min\{|\lambda|:\lambda\in\sigma(A)\}$ 当且仅当对应于模最小特征值的 A 的线性无关特征向量的个数等于该特征值的重数.

8. 设 $A,B\in\boldsymbol{M}_n(\mathbb{C})$，且有 $\boldsymbol{x}\neq\boldsymbol{0}$ 使得 $A\boldsymbol{x}=B\boldsymbol{x}$. 试证：对$\mathbb{C}^n$ 上任一有内点 $\boldsymbol{0}$ 的均衡凸体 S，$\text{glb}_S(B)\leqslant\text{lub}_S(A)$.

9. 设 $V\in\boldsymbol{M}_n(\mathbb{C})$为酉矩阵，试证：$\text{glb}_S(V)=\text{lub}_S(V)=1$，因而 $\sigma(V)$含在复平面单位圆周上.

2.3.3 广义矩阵范数与矩阵范数的补充

广义矩阵范数与矩阵范数理论有着丰富的内容. 在本小节仅补充其中的一部分知识. 首先介绍导出的矩阵范数的**极小性质**. 在实际中经常用条件 $\|A\|<1$ 来判定 $A\in\boldsymbol{M}_n(\mathbb{C})$的幂列$\{A^k\}_{k=1}^{\infty}$收敛于零矩阵(见 2.2.2 小节推论 8)，因此我们自然地喜欢那些一致地尽可能小的矩阵范数. 从后面讨论可以看出，$\boldsymbol{M}_n(\mathbb{C})$上导出的矩阵范数都具有这种极小的性质，反之亦然.

$\boldsymbol{M}_n(\mathbb{C})$上任意两个矩阵范数是等价的，因此，若 $\|\cdot\|_\alpha$ 与 $\|\cdot\|_\beta$ 为 $\boldsymbol{M}_n(\mathbb{C})$上矩阵

范数，则有正常数 $c_M(\alpha,\beta)$ 使得 $\|A\|_\alpha \leqslant c_M(\alpha,\beta)\|A\|_\beta$，$\forall A\in \boldsymbol{M}_n(\mathbb{C})$，这里 $c_M(\alpha,\beta)$ 可以取为

$$c_M(\alpha,\beta)=\max_{A\neq \mathbf{0}}\frac{\|A\|_\alpha}{\|A\|_\beta}. \tag{1}$$

交换 α 与 β 的地位，有 $c_M(\beta,\alpha)$ 使得 $\|A\|_\beta \leqslant c_M(\beta,\alpha)\|A\|_\alpha$，$\forall A\in \boldsymbol{M}_n(\mathbb{C})$. 这两个正常数 $c_M(\alpha,\beta)$ 与 $c_M(\beta,\alpha)$ 之间一般没有明显的关系. 但是，对于 $\|\cdot\|_1$（列范数），$\|\cdot\|_2$（谱范数）与 $\|\cdot\|_\infty$（行范数），我们有

$$\begin{aligned}
&\|A\|_1 \leqslant \sqrt{n}\,\|A\|_2, && \|A\|_2 \leqslant \sqrt{n}\,\|A\|_1,\\
&\|A\|_1 \leqslant n\,\|A\|_\infty, && \|A\|_\infty \leqslant n\,\|A\|_1,\\
&\|A\|_\infty \leqslant \sqrt{n}\,\|A\|_2, && \|A\|_2 \leqslant \sqrt{n}\,\|A\|_\infty.
\end{aligned} \tag{2}$$

因此，对于这三种特殊的导出的矩阵范数，均有 $c_M(\alpha,\beta)=c_M(\beta,\alpha)$（习题 1）. 下面定理表明，所有导出的矩阵范数都分享这种对称性质.

定理 1 设 $\|\cdot\|_\alpha$ 与 $\|\cdot\|_\beta$ 分别为 $\mathbb{C}^n$ 上向量范数 $\|\cdot\|$ 与 $\|\cdot\|'$ 导出的矩阵范数. 定义

$$R_{\alpha\beta}\equiv\max_{x\neq \mathbf{0}}\|\boldsymbol{x}\|/\|\boldsymbol{x}\|',R_{\beta\alpha}\equiv\max_{x\neq 0}\|\boldsymbol{x}\|'/\|\boldsymbol{x}\|. \tag{3}$$

则

$$\max_{A\neq O}\frac{\|A\|_\alpha}{\|A\|_\beta}=R_{\alpha\beta}R_{\beta\alpha}=\max_{A\neq O}\frac{\|A\|_\beta}{\|A\|_\alpha}. \tag{4}$$

证明 给定 $O\neq A\in \boldsymbol{M}_n(\mathbb{C})$ 与 $\mathbf{0}\neq \boldsymbol{x}\in\mathbb{C}^n$ 使得 $A\boldsymbol{x}\neq\mathbf{0}$. 这时有

$$\frac{\|A\boldsymbol{x}\|}{\|\boldsymbol{x}\|}=\frac{\|A\boldsymbol{x}\|}{\|A\boldsymbol{x}\|'}\frac{\|A\boldsymbol{x}\|'}{\|\boldsymbol{x}\|'}\frac{\|\boldsymbol{x}\|'}{\|\boldsymbol{x}\|}\leqslant R_{\alpha\beta}\frac{\|A\boldsymbol{x}\|'}{\|x\|'}R_{\beta\alpha}.$$

显然，当 $\boldsymbol{x}\neq\mathbf{0}$ 与 $A\boldsymbol{x}=\mathbf{0}$ 时，这个不等式（除去中间项）亦成立. 于是，我们有

$$\|A\|_\alpha\leqslant R_{\alpha\beta}R_{\beta\alpha}\|A\|_\beta,\qquad \forall A\in\boldsymbol{M}_n(\mathbb{C}),$$

因而

$$\frac{\|A\|_\alpha}{\|B\|_\beta}\leqslant R_{\alpha\beta}R_{\beta\alpha},\qquad \forall A\neq O.$$

另一方面，按(3)式有 $\boldsymbol{y},\boldsymbol{z}\in\mathbb{C}^n$ 使得 $\|\boldsymbol{y}\|_2=\|\boldsymbol{z}\|_2=1$，$\|\boldsymbol{y}\|=R_{\alpha\beta}\|\boldsymbol{y}\|'$ 与 $\|\boldsymbol{z}\|'=R_{\beta\alpha}\|\boldsymbol{z}\|$. 应用习题 2.3.1 题 5 的结果，存在 $\boldsymbol{z}_0\in\mathbb{C}^n$ 使得

$$|\boldsymbol{z}_0^*\boldsymbol{x}|\leqslant\|\boldsymbol{x}\|',\qquad \forall\boldsymbol{x}\in\mathbb{C}^n. \tag{5}$$

$$\boldsymbol{z}_0^*\boldsymbol{z}=\|\boldsymbol{z}\|'. \tag{6}$$

考虑矩阵 $A_0\equiv\boldsymbol{y}\boldsymbol{z}_0^*$. 按(6)式可推出

$$\frac{\|A_0\boldsymbol{z}\|}{\|\boldsymbol{z}\|}=\frac{\|\boldsymbol{y}\|\,|\boldsymbol{z}_0^*\boldsymbol{z}|}{\|\boldsymbol{z}\|}=\frac{\|\boldsymbol{y}\|\;\|\boldsymbol{z}\|'}{\|\boldsymbol{z}\|},$$

因而

$$\|A_0\|_\alpha\geqslant\frac{\|\boldsymbol{y}\|\;\|\boldsymbol{z}\|'}{\|\boldsymbol{z}\|}=R_{\alpha\beta}R_{\beta\alpha}\|\boldsymbol{y}\|'.$$

类似地按(5)式可得到

$$\frac{\| A_0 \boldsymbol{x} \|'}{\| \boldsymbol{x} \|'} = \frac{\| \boldsymbol{y} \|' \| z_0{}^* \boldsymbol{x} \|}{\| \boldsymbol{x} \|'} \leqslant \frac{\| \boldsymbol{y} \|' \| \boldsymbol{x} \|'}{\| \boldsymbol{x} \|'} = \| \boldsymbol{y} \|',$$

因而

$$\| A_0 \|_\beta \leqslant \| \boldsymbol{y} \|'.$$

于是，

$$\frac{\| A_0 \|_\alpha}{\| A_0 \|_\beta} \geqslant \frac{R_{\alpha\beta} R_{\beta\alpha} \| \boldsymbol{y} \|'}{\| \boldsymbol{y} \|'} = R_{\alpha\beta} R_{\beta\alpha}.$$

综上所证我们得到(4)式中第一个等式. 再应用此等式关于 α 与 β 的对称性立即推出(4)式的第二个等式. □

作为前一定理推论，我们有下面两个结果. 第一个表明，若不计正常数因子，$\mathbb{C}^n$ 上不同向量范数必对应着不同导出的矩阵范数. 第二个表明，不存在导出的矩阵范数一致地被另一个导出的矩阵范数占优.

推论 2 设 $\|\cdot\|$，$\|\cdot\|'$，$\|\cdot\|_\alpha$ 与 $\|\cdot\|_\beta$ 同定理 1. 则 $\| A \|_\alpha = \| A \|_\beta$，$\forall A \in \boldsymbol{M}_n(\mathbb{C})$，当且仅当存在正常数 c 使得 $\| \boldsymbol{x} \| = c \| \boldsymbol{x} \|'$，$\forall \boldsymbol{x} \in \mathbb{C}^n$.

推论 3 设 $\|\cdot\|$，$\|\cdot\|'$，$\|\cdot\|_\alpha$ 与 $\|\cdot\|_\beta$ 同定理 1. 则 $\| A \|_\alpha \leqslant \| A \|_\beta$，$\forall A \in \boldsymbol{M}_n(\mathbb{C})$，当且仅当 $\| A \|_\alpha = \| A \|_\beta$，$\forall A \in \boldsymbol{M}_n(\mathbb{C})$.

以上两个推论的证明留给读者(见习题 2,3).

如果将导出的矩阵范数同其他(不必导出的)矩阵范数相比较，则有如下极小性的结果.

定理 4 设 $\|\cdot\|$ 为 $\boldsymbol{M}_n(\mathbb{C})$ 任一给定的矩阵范数，$\|\cdot\|_\alpha$ 为 $\boldsymbol{M}_n(\mathbb{C})$ 上任一给定导出的矩阵范数. 则：

(1) 存在 $\boldsymbol{M}_n(\mathbb{C})$ 上导出的矩阵范数 $N(\cdot)$ 使得 $N(A) \leqslant \| A \|$，$\forall A \in \boldsymbol{M}_n(\mathbb{C})$.

(2) $\| A \| \leqslant \| A \|_\alpha$，$\forall A \in \boldsymbol{M}_n(\mathbb{C})$ 当且仅当 $\| A \| = \| A \|_\alpha$，$\forall A \in \boldsymbol{M}_n(\mathbb{C})$.

证明 (1) 对 $\boldsymbol{M}_n(\mathbb{C})$ 上给定的矩阵范数 $\|\cdot\|$，定义 $\mathbb{C}^n$ 上的向量范数 $\|\cdot\|'$ 如下：

$$\| \boldsymbol{x} \|' = \| X \|, X = [\boldsymbol{x} \cdots \boldsymbol{x}] \in \boldsymbol{M}_n(\mathbb{C}).$$

令 $N(\cdot)$ 为 $\boldsymbol{M}_n(\mathbb{C})$ 上 $\|\cdot\|'$ 导出的矩阵范数. 由于 $\|\cdot\|$ 为 $\boldsymbol{M}_n(\mathbb{C})$ 上矩阵范数，我们有

$$\begin{aligned} N(A) &= \sup_{x \neq 0} \frac{\| A\boldsymbol{x} \|'}{\| \boldsymbol{x} \|'} = \sup_{x \neq 0} \frac{\| [A\boldsymbol{x} \cdots A\boldsymbol{x}] \|}{\| [\boldsymbol{x} \cdots \boldsymbol{x}] \|} \\ &= \sup_{x \neq 0} \frac{\| AX \|}{\| X \|} \leqslant \| A \|, \qquad \forall A \in \boldsymbol{M}_n(\mathbb{C}). \end{aligned}$$

(2) 假设 $\| A \| \leqslant \| A \|_\alpha$，$\forall A \in \boldsymbol{M}_n(\mathbb{C})$，此时由(1)得到

$$N(A) \leqslant \| A \| \leqslant \| A \|_\alpha, \qquad \forall A \in \boldsymbol{M}_n(\mathbb{C}),$$

但按推论 3，上式应变为等式：$N(A)=\|A\|=\|A\|_\alpha$，$\forall A\in \boldsymbol{M}_n(\mathbb{C})$. □

假如 $\boldsymbol{M}_n(\mathbb{C})$ 上矩阵范数 $\|\cdot\|$ 满足条件：若 $\boldsymbol{M}_n(\mathbb{C})$ 上矩阵范数 $\|\cdot\|'$ 使得 $\|A\|'\leqslant\|A\|$，$\forall A\in \boldsymbol{M}_n(\mathbb{C})$，则 $\|\cdot\|'=\|\cdot\|$，那么便说 $\|\cdot\|$ 为**极小矩阵范数**(minimal matrix norm). 按此定义，定理 4(2)表明：$\boldsymbol{M}_n(\mathbb{C})$上每一个导出的矩阵范数为极小的，反之，由定理 4(1)直接推出，每一个极小矩阵范数为导出的矩阵范数.

下一个推论告诉我们，在 $\boldsymbol{M}_n(\mathbb{C})$上所有酉不变矩阵范数中只有一个极小矩阵范数，那就是谱范数 $\|\cdot\|_2$.

推论 5 若 $\|\cdot\|$ 为 $\boldsymbol{M}_n(\mathbb{C})$上任一酉不变矩阵范数，则 $\|A\|_2\leqslant\|A\|$，$\forall A\in \boldsymbol{M}_n(\mathbb{C})$. 并且，谱范数 $\|\cdot\|_2$ 是 $\boldsymbol{M}_n(\mathbb{C})$上唯一的既为导出的又为酉不变的矩阵范数.

证明 根据定理 4(1)，$N(A)\leqslant\|A\|$，$\forall A\in \boldsymbol{M}_n(\mathbb{C})$，这里 $N(A)$表示$\mathbb{C}^n$ 上向量范数 $\|\boldsymbol{x}\|'\equiv\|[\boldsymbol{x}\quad \boldsymbol{x}\quad\cdots\quad \boldsymbol{x}]\|$ 导出的矩阵范数，即 $N(A)=\sup\limits_{x\neq 0}\|A\boldsymbol{x}\|'/\|\boldsymbol{x}\|'=\sup\limits_{X\neq O}\|AX\|/\|X\|$，式中 $X\equiv[\boldsymbol{x}\quad \boldsymbol{x}\quad\cdots\quad \boldsymbol{x}]$为以 $\boldsymbol{x}$ 为列向量的 n 阶矩阵. 若 $U\in \boldsymbol{M}_n(\mathbb{C})$为酉矩阵，则 $\|U\boldsymbol{x}\|'=\|UX\|=\|X\|=\|\boldsymbol{x}\|'$，因而 $\|\cdot\|'$为$\mathbb{C}^n$ 上酉不变的向量范数. 现设 $\boldsymbol{x}\neq\boldsymbol{0}$ 为$\mathbb{C}^n$ 的任意给定的向量，则有酉矩阵 $U\in \boldsymbol{M}_n(\mathbb{C})$使得 $U\boldsymbol{x}=\|\boldsymbol{x}\|_2\boldsymbol{e}_1$，于是 $\|\boldsymbol{x}\|'=\|\,\|\boldsymbol{x}\|_2U^*\boldsymbol{e}_1\|'=\|\boldsymbol{x}\|_2\|\boldsymbol{e}_1\|'$，$\forall \boldsymbol{x}\in\mathbb{C}^n$. 按推论 2，$N(\cdot)$等于 $\boldsymbol{M}_n(\mathbb{C})$上矩阵谱范数 $\|\cdot\|_2$，因而 $\|A\|_2=N(A)\leqslant\|A\|$，$\forall A\in \boldsymbol{M}_n(\mathbb{C})$. 假如 $\|\cdot\|$ 还是导出的矩阵范数，那么它是极小的(见定理 4(2))，因而 $\|A\|_2=\|A\|$，$\forall A\in \boldsymbol{M}_n(\mathbb{C})$. □

现在介绍**自伴矩阵范数**. 对 $\boldsymbol{M}_n(\mathbb{C})$上任一矩阵范数 $\|\cdot\|$，定义实值函数 $\|\cdot\|^*$：

$$\|A\|^*=\|A^*\|,\qquad \forall A\in \boldsymbol{M}_n(\mathbb{C}).\tag{7}$$

容易看出. $\|\cdot\|^*$ 也是 $\boldsymbol{M}_n(\mathbb{C})$上的矩阵范数，并且，$\|\cdot\|^{**}\equiv(\|\cdot\|^*)^*=\|\cdot\|$. 特殊地，我们有 $\|A\|_2^*=\|A^*\|_2=\|A\|_2$ 与 $\|A\|_1^*=\|A^*\|_1=\|A\|_\infty\neq\|A\|_1$，$\|A\|_\infty^*=\|A^*\|_\infty=\|A\|_1\neq\|A\|_\infty$，$\|A\|_F^*=\|A^*\|_F=\|A\|_F$. 通常将满足条件 $\|\cdot\|^*=\|\cdot\|$ 的 $\boldsymbol{M}_n(\mathbb{C})$上的矩阵范数称为**自伴的**(self-adjoint). 因此，$\boldsymbol{M}_n(\mathbb{C})$上的谱范数与 Frobenius 范数为自伴的，行范数与列范数不是自伴的. 一般地，所有 $\boldsymbol{M}_n(\mathbb{C})$上的酉不变矩阵范数皆为自伴的，我们略去它的证明，详情见文献[5].

$\boldsymbol{M}_n(\mathbb{C})$上的谱范数是 $\boldsymbol{M}_n(\mathbb{C})$上唯一的既是导出的又是自伴的矩阵范数，这个结论构成下列定理的一部分.

定理 6 设 $\|\cdot\|$ 是任意给定的 $\boldsymbol{M}_n(\mathbb{C})$上矩阵范数. 则：

(1) $\|\cdot\|^*$ 为导出的矩阵范数当且仅当 $\|\cdot\|$ 为导出的矩阵范数.

(2) 若 $\|\cdot\|$ 为$\mathbb{C}^n$ 上向量范数 $\|\cdot\|_\alpha$ 导出的矩阵范数，则 $\|\cdot\|^*$ 为 $\|\cdot\|_\alpha$的对偶范

数$(\|\cdot\|_\alpha)_D$导出的矩阵范数.

(3) 谱范数$\|\cdot\|_2$是$\boldsymbol{M}_n(\mathbb{C})$上唯一的既是导出的又是自伴的矩阵范数.

证明 若$N(\cdot)$为$\boldsymbol{M}_n(\mathbb{C})$上矩阵范数满足$N(A)\leqslant\|A\|^*=\|A^*\|$，$\forall A\in\boldsymbol{M}_n(\mathbb{C})$，则$N(A)^*\equiv N(A^*)\leqslant\|A\|$，$\forall A\in\boldsymbol{M}_n(\mathbb{C})$. 若$\|\cdot\|$为极小的，则有$N(\cdot)^*=\|\cdot\|$，因而$(N(\cdot)^*=N(\cdot)=\|\cdot\|^*$，这表明$\|\cdot\|^*$为极小的矩阵范数. 但$\boldsymbol{M}_n(\mathbb{C})$上矩阵范数是导出的等价于它是极小的(见定理 4 或习题 4)，因而(1)得证. 现在证明(2). 设$\|\cdot\|$为$\mathbb{C}^n$上向量范数$\|\cdot\|_\alpha$导出的矩阵范数，应用对偶性定理(2.3.1 小节定理 5)，则我们有

$$\begin{aligned}\|A\|^*&=\|A^*\|=\max_{\|x\|_\alpha=1}\|A^*x\|_\alpha=\max_{\|x\|_\alpha=1}(\|A^*x\|_\alpha)_{DD}\\&=\max_{\|x\|_\alpha=1}\max_{(\|z\|_\alpha)_D=1}|(A^*x)^*z|=\max_{\|z\|_{\alpha D}=1}\max_{\|x\|_\alpha=1}|x^*Az|\\&=\max_{(\|z\|_\alpha)_D=1}(\|Az\|_\alpha)_D,\end{aligned}$$

因而$\|\cdot\|^*$为$(\|\cdot\|_\alpha)_D$导出的矩阵范数. 为证(3)，我们看到如$\|\cdot\|$为$\mathbb{C}^n$上向量范数$\|\cdot\|_\alpha$导出的矩阵范数，且$\|\cdot\|=\|\cdot\|^*$，则(2)断言$\|\cdot\|$也为$(\|\cdot\|_\alpha)_D$导出的矩阵范数. 因此，按推论 2，存在正数c使得$(\|\cdot\|_\alpha)_D=c\|\cdot\|_\alpha$. 但由 2.3.1 小节定理 3 这时有$\|\cdot\|_\alpha=\|\cdot\|_2/\sqrt{c}$. 由于$\|\cdot\|_\alpha$为$\mathbb{C}^n$上欧氏范数$\|\cdot\|_2$的正倍数，故按推论 2 它们有相同的导出的矩阵范数，即有$\|A\|=\|A\|_2$，$\forall A\in\boldsymbol{M}_n(\mathbb{C})$. □

本小节余下篇幅将用以介绍矩阵$A\in\boldsymbol{M}_n(\mathbb{C})$的**数值域**(numerical range)与**数值半径**(numerical radius)的概念.

定义 7 若$A\in\boldsymbol{M}_n(\mathbb{C})$为给定的矩阵，则$A$的**数值域**为复平面$\mathbb{C}$内区域$W(A)$：

$$W(A)\equiv\{x^*Ax:x\in\mathbb{C}^n\text{ 与 }x^*x=1\};\tag{8}$$

A的**数值半径**$r(A)$为

$$r(A)\equiv\max_{x^*x=1}|x^*Ax|=\max_{z\in W(A)}|z|.\tag{9}$$

按照(8)式，$W(A)$为定义在$\mathbb{C}^n$内单位欧氏球面上的下列映射的值域：

$$x\rightarrow x^*Ax.$$

取x为A的特征向量，满足$x^*x=1$，则由(8)式看出，A的谱$\sigma(A)\subset W(A)$，因而$\rho(A)\leqslant r(A)$. 此外，由(9)式看出，$r(A)=r(A^*)$，$\forall A\in\boldsymbol{M}_n(\mathbb{C})$，并且，$r(A)=r(U^*AU)$，这里$U\in\boldsymbol{M}_n(\mathbb{C})$为任意的酉矩阵. 当$A\in\boldsymbol{M}_n(\mathbb{C})$为正规矩阵时，因$A$为酉可对角化的，所以有$r(A)=\rho(A)=\|A\|_2$. 但对一般$A\in\boldsymbol{M}_n(\mathbb{C})$，只有$r(A)\leqslant\max_{\|x\|_2=1}\|Ax\|_2=\|A\|_2$，因为$|x^*Ax|\leqslant\|Ax\|_2\|x\|_2$，$\forall x\in\mathbb{C}^n$. 例如考虑

$$A=\begin{bmatrix}0&0\\1&0\end{bmatrix},$$

$W(A)$为复平面上圆心在原点半径为 1/2 的闭圆，因而$r(A)=1/2$，但$\|A\|_2=$

$(\rho(AA^*))^{1/2}=1$. 对于 $r(A)$ 与 $\|A\|_2$,可以证明更好的结论:

$$\frac{1}{2}\|A\|_2 \leqslant r(A) \leqslant \|A\|_2, \qquad \forall A \in \boldsymbol{M}_n(\mathbb{C}). \tag{10}$$

事实上对此只需证明: $\|A\|_2 \leqslant 2r(A)$, $\forall A \in \boldsymbol{M}_n(\mathbb{C})$. 将 A 写成

$$A = \frac{A+A^*}{2} + \frac{A-A^*}{2} \equiv A_1 + A_2,$$

则 A_1 与 A_2 均为正规的. 于是,

$$\|A\|_2 \leqslant \|A_1\|_2 + \|A_2\|_2 = r(A_1) + r(A_2) \leqslant r(A) + r(A^*) = 2r(A).$$

作为(10)式的简单推论,我们有:若 $A \in \boldsymbol{M}_n(\mathbb{C})$,则

$$A = O \Leftrightarrow W(A) = \{0\}. \tag{11}$$

矩阵 A 的数值域的一个基本结果是如下的 **Toeplitz-Hausdorff 定理**(见文献[6]).

定理 8 设 $A \in \boldsymbol{M}_n(\mathbb{C})$. 则 $W(A)$ 为复平面 $\mathbb{C}$ 内的凸集.

证明 任取 $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^n$,使得 $\boldsymbol{x}^* \boldsymbol{x} = \boldsymbol{y}^* \boldsymbol{y} = 1$,并令 $\xi = \boldsymbol{x}^* A \boldsymbol{x}$ 与 $\eta = \boldsymbol{y}^* A \boldsymbol{y}$. 我们要证: $\lambda\xi + (1-\lambda)\eta \in W(A)$, $0 \leqslant \lambda \leqslant 1$. 若 $\xi = \eta$, 结论显然成立. 若 $\xi \neq \eta$,则有 $\alpha, \beta \in \mathbb{C}$ 使得 $\alpha\xi + \beta = 1$ 与 $\alpha\eta + \beta = 0$. 这时只需证明: $W(\alpha A + \beta I)(=\alpha W(A) + \beta)$ 包含复平面内闭区间 $[0,1]$. 因为如令 $t = \alpha \boldsymbol{z}^* A \boldsymbol{z} + \beta \in [0,1]$,那么,

$$\alpha \boldsymbol{z}^* A \boldsymbol{z} + \beta = t(\alpha\xi + \beta) + (1-t)(\alpha\eta + \beta) = \alpha(t\xi + (1-t)\eta) + \beta.$$

因此,不失普遍性,我们可以假定 $\xi = \boldsymbol{x}^* A \boldsymbol{x} = 1$ 与 $\eta = \boldsymbol{y}^* A \boldsymbol{y} = 0$. 这时 $\alpha = 1, \beta = 0$,即只要证明: $[0,1] \subset W(A)$.

现将 A 写成 $A = B + \mathrm{i}C$,其中 $B = \mathrm{Re}A$ 与 $C = \mathrm{Im}A$ 均为 Hermite 矩阵. 由于 $\boldsymbol{x}^* A \boldsymbol{x} = 1$ 与 $\boldsymbol{y}^* A \boldsymbol{y} = 0$,故 $\boldsymbol{x}^* C \boldsymbol{x} = \boldsymbol{y}^* C \boldsymbol{y} = 0$. 若以 $\lambda \boldsymbol{x}(|\lambda| = 1)$ 代替 $\boldsymbol{x}$,则 $\boldsymbol{x}^* A \boldsymbol{x}$ 保持不变,但 $\boldsymbol{y}^* C \boldsymbol{x}$ 变为 $\lambda \boldsymbol{y}^* C \boldsymbol{x}$. 因此,不失普遍性,可以假定: $\boldsymbol{y}^* C \boldsymbol{x}$ 为纯虚数.

有了这些约定后,置 $\boldsymbol{z}(t) = t\boldsymbol{x} + (1-t)\boldsymbol{y}$, $0 \leqslant t \leqslant 1$. 首先有 $\boldsymbol{z}(t) \neq \boldsymbol{0}$. 这是因为 $\boldsymbol{x}^* A \boldsymbol{x} \neq \boldsymbol{y}^* A \boldsymbol{y}$,若 $\boldsymbol{z}(t) = \boldsymbol{0}$,则 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 线性相关,于是, $\boldsymbol{y} = \lambda \boldsymbol{x}$, $|\lambda| = 1$(因为 $\boldsymbol{x}^* \boldsymbol{x} = \boldsymbol{y}^* \boldsymbol{y} = 1$),因而 $\boldsymbol{x}^* A \boldsymbol{x} = \boldsymbol{y}^* A \boldsymbol{y}$. 由于

$$\begin{aligned} \boldsymbol{z}(t)^* C \boldsymbol{z}(t) &= t^2 \boldsymbol{x}^* C \boldsymbol{x} + t(1-t)(\boldsymbol{y}^* C \boldsymbol{x} + \overline{\boldsymbol{y}^* C \boldsymbol{x}}) + (1-t)^2 \boldsymbol{y}^* C \boldsymbol{y} \\ &= t(1-t)\mathrm{Re}\boldsymbol{y}^* C \boldsymbol{x} = 0, \qquad \forall t \in [0,1], \end{aligned}$$

故 $\boldsymbol{z}(t)^* A \boldsymbol{z}(t)$ 对所有 $t \in [0,1]$ 为实数. 另一方面,如下定义的实值函数:

$$t \to \boldsymbol{z}(t)^* A \boldsymbol{z}(t) / [\boldsymbol{z}(t)^* \boldsymbol{z}(t)]$$

在 $t \in [0,1]$ 上连续,且当 $t = 0$ 与 $t = 1$ 时其值分别为 0 与 1. 因此,此函数的值域包含闭区间 $[0,1]$,因而 $W(A)$ 包含闭区间 $[0,1]$. □

我们感兴趣的问题是：$A\in \boldsymbol{M}_n(\mathbb{C})$的数值半径 $r(A)$是不是 $\boldsymbol{M}_n(\mathbb{C})$上的广义矩阵范数或矩阵范数？由下面例子

$$A=\begin{bmatrix}0 & 0\\ 1 & 0\end{bmatrix} \quad 与 \quad B=\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}$$

看出，$r(A)=r(B)=1/2, r(AB)=1$，因而 $r(\cdot)$不是 $\boldsymbol{M}_n(\mathbb{C})$上的矩阵范数. 但是，我们有

定理 9　数值半径 $r(\cdot)$是 $\boldsymbol{M}_n(\mathbb{C})$上的广义矩阵范数.

证明　非负性、齐次性与三角不等式条件成立都是明显的. 而正性等价于(11)式. □

习题 2.3.3

1. 验证不等式(2).

2. 证明推论 2.

3. 证明推论 3.

4. 设 $\|\cdot\|$ 为 $\boldsymbol{M}_n(\mathbb{C})$上矩阵范数. 若 $\boldsymbol{y}\neq\boldsymbol{0}$ 为$\mathbb{C}^n$ 内的给定向量，定义 $\|\boldsymbol{x}\|_y=\|\boldsymbol{x}\boldsymbol{y}^*\|, \forall \boldsymbol{x}\in\mathbb{C}^n$. 试证：$\|\cdot\|_y$ 为$\mathbb{C}^n$ 上向量范数，它满足 $\|A\boldsymbol{x}\|_y\leqslant\|A\|\ \|\boldsymbol{x}\|_y, \forall A\in\boldsymbol{M}_n(\mathbb{C})$与 $\forall \boldsymbol{x}\in\mathbb{C}^n$. 此外，设 $N_y(\cdot)$为 $\|\cdot\|_y$ 导出的矩阵范数，试证下列诸命题等价：

(1) $\|\cdot\|$ 为导出的矩阵范数.

(2) $\|\cdot\|$ 为极小的矩阵范数.

(3) $\|\cdot\|=N_y(\cdot), \forall \boldsymbol{0}\neq\boldsymbol{y}\in\mathbb{C}^n$.

5. 设 $\|\cdot\|$ 为$\mathbb{C}^n$ 上向量范数 $\|\cdot\|_\alpha$ 导出的矩阵范数. 试证：$\|\boldsymbol{y}\boldsymbol{z}^*\|=\|\boldsymbol{y}\|_\alpha(\|\boldsymbol{z}\|_\alpha)_D, \|\cdot\|_z=\|\cdot\|_\alpha(\|\boldsymbol{z}\|_\alpha)_D$ 与 $\|\cdot\|_y=\|\cdot\|_z(\|\boldsymbol{y}\|_\alpha)_D/(\|\boldsymbol{z}\|_\alpha)_D, \forall \boldsymbol{y},\boldsymbol{z}\in\mathbb{C}^n$，这里，$\|\cdot\|_y$ 与 $\|\cdot\|_z$ 的定义见前题，$(\|\cdot\|_\alpha)_D$ 为 $\|\cdot\|_\alpha$ 的对偶范数.

6. 设 $\|\cdot\|$，$\|\cdot\|_y$ 意义同题 4. 试证下列命题等价：

(1) 对$\mathbb{C}^n$ 内每对非零向量 $\boldsymbol{y}$ 与 $\boldsymbol{z}$ 存在正常数 c_{yz} 使得 $\|\boldsymbol{x}\|_y=c_{yz}\|\boldsymbol{x}\|_z, \forall \boldsymbol{x}\in\mathbb{C}^n$.

(2) $\|\boldsymbol{x}\boldsymbol{y}^*\|=\|\boldsymbol{x}\boldsymbol{z}^*\|\|\boldsymbol{z}\boldsymbol{y}^*\|/\|\boldsymbol{z}\boldsymbol{z}^*\|, \forall \boldsymbol{x},\boldsymbol{y},\boldsymbol{z}\in\mathbb{C}^n, \boldsymbol{z}\neq\boldsymbol{0}$.

此外，若 $\|\cdot\|$ 还是 $\boldsymbol{M}_n(\mathbb{C})$上导出的矩阵范数，则它满足(2)中的恒等式，并且由它构造出来的向量范数 $\|\cdot\|_y$ 满足(1).

7. 若

$$A=\begin{bmatrix}0 & 0\\ 1 & 1\end{bmatrix},$$

求 $W(A)$与 $r(A)$.

8. 试证：对任意 $A\in\boldsymbol{M}_n(\mathbb{C})$，$W(A)$为$\mathbb{C}$ 中有界闭集.

9. 证明：当且仅当 A 为 Hermite 矩阵时，$W(A)$为实轴上的一个区间.

10. 试证：若 $r(I_n-A)<1$，则 $A\in\boldsymbol{M}_n(\mathbb{C})$非奇异；若 $r(A)=\|A\|_2$，则 $\rho(A)=\|A\|_2$.

11. 设 $A\in\boldsymbol{M}_n(\mathbb{C})$使得 $r(A)\leqslant 1$，试证：对所有正整数 k，$r(A^k)\leqslant 1$.

参考文献

[1] Householder A S. The Theory of Matrices in Numerical Analysis. New York: Blaisdell, 1964, Reprint by Dover, New York, 1975(前三章有中译本,孙家昶等译. 数值分析中的矩阵论. 北京:科学出版社,1986)

[2] Horn R A, Johnson. C R. Matrix Analysis. Cambridge: Cambridge Univ. Press, 1994

[3] Lancaster P, Tismenetsky M. The Theory of Matrices with Applications. 2nd ed.. New York: Academic Press, 1985

[4] Fiedler M. Special Matrices and their Applications in Numerical Mathematics. Boston, Martinus Nijhoff Pub, 1986

[5] von Neumann J. Some matrix-inequalities and metrization of matrix-space, Bull. Inst. Math. Mécan. Univ. Kouybycheff Tomsk, 1(1935～1937), 286～300

[6] Halmos P R. A Hilbert Space Problem Book. New York: Van Nostrand, 1967: 166

[7] Stewart G W. Introduction to Matrix Computations. New York: Academic Press, 1973(中译本,王国荣等译. 矩阵计算引论. 上海:上海科学技术出版社,1980)

[8] Collatz L. Functional Analysis and Numerical Mathematics. New York: Academic Press, 1966

[9] Marcus M. Topics in Linear Algebra. Monograph Series of Institute for the Interdisip linary Applications of Algebra and Combinatorics. UC, Santa, Barbara, 1977.

[10] Glazman I M, Liubich J I. Finite-Dimensional linear Analysis. Cambridge, Mass: M. I. T. Press, 1974

[11] Golub G H, Van Loan C F. Matrix Computations. 3rd ed. Baltimore and London: Johns Hopkins Univ. Press, 1997

[12] Rudin W. Functional Analysis. New York: McGraw-Hill, Inc. , 1973

第三章 矩阵函数

我们已经接触到关于矩阵的一些特殊函数. 譬如,Cayley-Hamilton 定理(见第一章 1.1.2 小节定理 3)指出,任意矩阵 $A\in \boldsymbol{M}_n(\mathbb{C})$满足自身的特征多项式,换句话说,若

$$f(\lambda)\equiv \det(\lambda I_n-A)=\lambda^n+a_{n-1}\lambda^{n-1}+\cdots+a_1\lambda+a_0,$$

则有

$$f(A)=A^n+a_{n-1}A^{n-1}+\cdots+a_1A+a_0I_n=0.$$

由上式确定的函数 $f(A)$便是从 $\boldsymbol{M}_n(\mathbb{C})$到 $\boldsymbol{M}_n(\mathbb{C})$内的矩阵多项式函数. 一般地,如给定复系数多项式 $p(\lambda)=\sum\limits_{j=0}^{m}\alpha_j\lambda^j$,自然地可定义一个对应的矩阵多项式函数 $p(A)=\sum\limits_{j=0}^{m}\alpha_jA^j,\forall A\in \boldsymbol{M}_n(\mathbb{C})$. 映射 $p(\lambda)\to p(A)$ 为多项式代数到矩阵代数 $\boldsymbol{M}_n(\mathbb{C})$ 内的同态,这意味着 $p(\lambda)+q(\lambda)=r(\lambda)$ 与 $p(\lambda)q(\lambda)=s(\lambda)$ 蕴涵 $p(A)+q(A)=r(A)$ 与 $p(A)q(A)=s(A)$. 由此特别推出 $p(A)q(A)=q(A)p(A)$. 但问题不总是这样简单的. 例如,给定 $f(\lambda)=1/(\alpha-\lambda)$ 与 $A\in \boldsymbol{M}_n(\mathbb{C})$,为了使 $f(A)=(\alpha I_n-A)^{-1}$ 有意义,显然必须要求 $\alpha\notin\sigma(A)$,这里 $\sigma(A)$ 为 A 的谱.

在本章我们系统地讨论矩阵函数理论. 对给定函数 $f:\mathbb{C}\to\mathbb{C}$ 与矩阵 $A\in \boldsymbol{M}_n(\mathbb{C})$,讨论 $f(A)$的涵义、有关性质与应用. 我们指出,本章的基本结果都可以推广到无穷维 Hilbert 空间上有界线性算子的情形(见文献[2]).

3.1 简单矩阵的函数

3.1.1 定义

设 A 为 $\boldsymbol{M}_n(\mathbb{C})$内的简单矩阵,即为可对角化的,按定义存在非奇异 $P\in \boldsymbol{M}_n(\mathbb{C})$使得

$$A=PDP^{-1},\tag{1}$$

式中,$D=\mathrm{diag}[\lambda_1,\lambda_2,\cdots,\lambda_n]$,$\lambda_j\in\sigma(A)$. (1)式推出,$A^k=PD^kP^{-1}$,$k=0,1,2,\cdots$,因而对任意的数值多项式 $p(\lambda)$有 $p(A)=Pp(D)P^{-1}$,式中 $p(D)=\mathrm{diag}[p(\lambda_1),\cdots,p(\lambda_n)]$. 一般地,对给定的简单矩阵 $A\in \boldsymbol{M}_n(\mathbb{C})$,它有表达式(1),假如单复变量 λ 的复值函数 f 有值 $f(\lambda_1),\cdots,f(\lambda_n)$,则 f 叫作**在 $\sigma(A)$ 上有定义**. 受多项式情形的启发,我们引出如下定义.

定义 1　设 $A\in \boldsymbol{M}_n(\mathbb{C})$ 有形式(1). 若复值函数 f 在 $\sigma(A)$ 上有定义，则 $f(A)$ 定义为

$$f(A) = P\text{diag}[f(\lambda_1),\cdots,f(\lambda_n)]P^{-1}. \tag{2}$$

由此定义我们有如下结论.

定理 2(谱映射定理)　设 $A\in \boldsymbol{M}_n(\mathbb{C})$ 为简单矩阵，f 在 $\sigma(A)$ 上有定义. 则 $\sigma(f(A))=f(\sigma(A))$.

从(2)式看出，$f(A)$ 只与 $f(\lambda)$ **在 $\boldsymbol{\sigma}(\boldsymbol{A})$ 上的值** $f(\lambda_1),\cdots,f(\lambda_n)$ 有关(由后面定理 3 看出，它与相似变换矩阵 P 的选取无关)，即若另有复值函数 $g(\lambda)$ 满足 $g(\lambda_i)=f(\lambda_i),i=1,\cdots,n$，则有 $g(A)=f(A)=P\text{diag}[g(\lambda_1),\cdots,g(\lambda_n)]P^{-1}$. 这样的 g 显然有无穷多个，我们希望选取尽可能简单的函数，例如多项式 $p(\lambda)$，使得 $p(\lambda_j)=f(\lambda_j),j=1,\cdots,n$. 这时必定有 $p(A)=f(A)$.

应用复系数多项式插值知识容易看出这样的多项式 $p(\lambda)$ 的存在性. 若适当限制范围，还可以推出这样多项式的唯一性. 具体如下.

设 A 的特征值 $\lambda_1,\cdots,\lambda_n$ 有 s 个不同值，不妨设为前 s 个：$\lambda_1,\lambda_2,\cdots,\lambda_s$. 记 $f_j=f(\lambda_j),j=1,\cdots,s$. 则在 $\mathscr{P}_{s-1}$(次数不超过 $s-1$ 的复系数多项式的线性空间)内存在唯一多项式 $p(\lambda)$ 满足 $p(\lambda_j)=f_j,j=1,\cdots,s$. 这个多项式 $p(\lambda)$ 可用 **Lagrange 公式**表示：

$$p(\lambda) = \sum_{k=1}^{s} f_k\varphi_k(\lambda), \tag{3}$$

式中，

$$\varphi_k(\lambda) \equiv \prod_{\substack{j=1\\ j\neq k}}^{s}(\lambda-\lambda_j)\Big/\prod_{\substack{j=1\\ j\neq k}}^{s}(\lambda_k-\lambda_j),\quad k=1,\cdots,s \tag{4}$$

称为这个插值问题的**基函数**，它们满足

$$\varphi_k(\lambda_j) = \delta_{kj},\qquad 1\leqslant k,j\leqslant s, \tag{5}$$

且仅依赖于 $\lambda_1,\cdots,\lambda_s$ 而与 $f_1,\cdots,f_s$ 无关. 我们指出，由(4)式确定的 s 个函数组成 $\mathscr{P}_{s-1}$ 的基. 事实上，$\dim\mathscr{P}_{s-1}=s$，且若 $\sum_{k=1}^{s}c_k\varphi_k(\lambda)=0$，则 $\sum_{k=1}^{s}c_k\varphi_k(\lambda_j)=\sum_{k=1}^{s}c_k\delta_{kj}=c_j=0,j=1,\cdots,s$. 因此，$\varphi_1,\cdots,\varphi_s$ 线性无关.

于是，我们有如下基本的结果.

定理 3　设 $A\in \boldsymbol{M}_n(\mathbb{C})$ 为简单矩阵，它有形式(1)，而 f 为在 $\sigma(A)$ 上有定义的复值函数. 若 A 的所有不同特征值为 $\lambda_1,\cdots,\lambda_s$，则有 $p(\lambda)\in\mathscr{P}_{s-1}$ 使得 $p(\lambda_j)=f(\lambda_j),j=1,\cdots,s$，与

$$p(A) = f(A), \tag{6}$$

并且，$f(A)$ 与(1)式中的 P 的选取无关，只与 $f(\lambda_1),\cdots,f(\lambda_s)$ 有关.

例 4 设 $f(\lambda)=e^{2\lambda}$ 与 $g(\lambda)=\lambda^{1/2}$. 对于

$$A=\begin{bmatrix}6 & -\sqrt{5}\\ -\sqrt{5} & 2\end{bmatrix}$$

求 $f(A)$ 与 $g(A)$.

这时 A 为实对称正定矩阵, $\sigma(A)=\{1,7\}$. 对应的公式(3)与(4)为

$$p(\lambda)=f(1)\frac{\lambda-7}{1-7}+f(7)\frac{\lambda-1}{7-1}=-\frac{1}{6}e^{2}(\lambda-7)+\frac{1}{6}e^{14}(\lambda-1). \tag{7}$$

因此,按定理 3,

$$\begin{aligned}f(A)=e^{2A}&=-\frac{1}{6}e^{2}(A-7I_2)+\frac{1}{6}e^{14}(A-I_2)\\&=\frac{1}{6}\begin{bmatrix}5e^{14}+e^{2} & -\sqrt{5}e^{14}+\sqrt{5}e^{2}\\ -\sqrt{5}e^{14}+\sqrt{5}e^{2} & e^{14}+5e^{2}\end{bmatrix}.\end{aligned}$$

对于 $g(\lambda)$, 选取 $\lambda^{1/2}$ 的主分枝, 因而 $g(1)=1, g(7)=\sqrt{7}$. 将(7)式中 $f(1)$ 与 $f(7)$ 分别换为 $g(1)$ 与 $g(7)$, 便得

$$\begin{aligned}g(A)\equiv A^{1/2}&=-\frac{1}{6}(A-7I_2)+\frac{1}{6}\sqrt{7}(A-I_2)\\&=\frac{1}{6}\begin{bmatrix}1+5\sqrt{7} & \sqrt{5}(1-\sqrt{7})\\ \sqrt{5}(1-\sqrt{7}) & 5+\sqrt{7}\end{bmatrix}.\end{aligned}$$

容易验证, e^{2A} 与 $A^{1/2}$ 均为实对称正定矩阵, 并且, $(A^{1/2})^2=A$, 即 $A^{1/2}$ 为 A 的对称正定平方根矩阵.

若取 $\lambda^{1/2}$ 的另一单值分枝使得 $g(1)=-1$ 与 $g(7)=-\sqrt{7}$, 则得 $g(A)=-A^{1/2}$, 它也是 A 的平方根, 因为 $(g(A))^2=A$. □

对简单矩阵 $A\in \boldsymbol{M}_n(\mathbb{C})$, 今后用记号 $\mathscr{H}$ 表示定义在 $\sigma(A)$ 上复值函数的集合. 若 $f,g\in\mathscr{H},\alpha\in\mathbb{C}$, 定义 $(f+g)(\lambda)=f(\lambda)+g(\lambda)$, $(\alpha f)(\lambda)=\alpha f(\lambda)$ 与 $(f\cdot g)(\lambda)=f(\lambda)g(\lambda)$, 则 $f+g,\alpha f,f\cdot g\in\mathscr{H}$. 容易看出, 对于加法与数乘运算, $\mathscr{H}$ 组成复向量空间; 对于乘法运算, 此向量空间组成一个可交换复代数. 按照 $f(A)$ 的定义, 我们有

$$\begin{aligned}(\alpha f+\beta g)(A)&=\alpha f(A)+\beta g(A), \quad \forall \alpha,\beta\in\mathbb{C},\\(f\cdot g)(A)&=f(A)g(A)=g(A)f(A).\end{aligned} \tag{8}$$

这表明: 当给定简单矩阵 $A\in\boldsymbol{M}_n(\mathbb{C})$ 时, 映射 φ:

$$\varphi(f)=f(A), \quad \forall f\in\mathscr{H} \tag{9}$$

为由 $\mathscr{H}$ 到矩阵代数 $\boldsymbol{M}_n(\mathbb{C})$ 内的一个**同态**. 它将 1 映射到 I_n, $f(\lambda)=\lambda$ 映身到 A; 若 $\det A\neq 0$, 则它将 $f(\lambda)=\lambda^{-1}$ 映射到 A^{-1}. 特殊地, 若 $A\in\boldsymbol{M}_n(\mathbb{C})$ 为正规矩阵(这时(1)式中 P 可取为酉矩阵 $U\in\boldsymbol{M}_n(\mathbb{C})$), 则 φ 将 $f(\lambda)=\bar{\lambda}$ 映射到 A^*.

习题 3.1.1

1. 设$A\in \boldsymbol{M}_n(\mathbb{C})$的特征多项式为

$$f(\lambda)=\det(\lambda I_n-A)=\lambda^n+a_{n-1}\lambda^{n-1}+\cdots+a_1\lambda+a_0.$$

试证：当$a_0\neq 0$时，A非奇异且A^{-1}的特征多项式为

$$g(\lambda)=\lambda^n+\frac{a_1}{a_0}\lambda^{n-1}+\cdots+\frac{a_{n-1}}{a_0}\lambda+\frac{1}{a_0}=\frac{1}{a_0}\lambda^n f(\lambda^{-1}).$$

2. 设$A=\operatorname{diag}[A_1,A_2,\cdots,A_m]$，其中各$A_j$为简单矩阵. 试证：若$f$为定义在$\sigma(A)$上的复值函数，则

$$f(A)=\operatorname{diag}[f(A_1),f(A_2),\cdots,f(A_m)].$$

3. 设$f(\lambda)=e^{3\lambda}$与$g(\lambda)=\sin\lambda$，而A为例4中的二阶矩阵，求$f(A)$与$g(A)$.

3.1.2 简单矩阵函数的谱分解及其应用

由3.1.1小节定理3与$p(\lambda)$的Lagrange公式，我们得到

$$f(A)=p(A)=\sum_{k=1}^{s}f(\lambda_k)\varphi_k(A),$$

$$\varphi_k(A)=\prod_{\substack{j=1\\j\neq k}}^{s}(A-\lambda_jI_n)\Big/\prod_{\substack{j=1\\j\neq k}}^{s}(\lambda_k-\lambda_j),\qquad k=1,\cdots,s. \tag{1}$$

这里，$\varphi_k(A)(k=1,\cdots,s)$称为$A$的**分量（矩阵）**，它们只与$A\in\boldsymbol{M}_n(\mathbb{C})$的$s$个不同特征值$\lambda_1,\cdots,\lambda_s$有关. 并且，$\varphi_1(A),\cdots,\varphi_s(A)$线性无关，因为若$\sum_{j=1}^{s}c_j\varphi_j(A)=O$，则$g(\lambda)\equiv\sum_{j=1}^{s}c_j\varphi_j(\lambda)$满足$g(A)=O$，因而$g(\lambda)$为$A$的零化多项式或零函数，但由于$g\in\mathscr{P}_{s-1}$，而$A$的零化多项式次数不低于$s$，故有$g(\lambda)\equiv 0$，再由$\varphi_1(\lambda),\cdots,\varphi_s(\lambda)$的线性无关性立即得出$c_1=c_2=\cdots=c_s=0$.

按3.1.1小节定义1，$A\in\boldsymbol{M}_n(\mathbb{C})$的分量$\varphi_k(A)$可以表示为另一形式：

$$\varphi_k(A)=P\operatorname{diag}[\varphi_k(\lambda_1),\cdots,\varphi_k(\lambda_n)]P^{-1}=P\operatorname{diag}[0,\cdots,0,1,\cdots,1,0,\cdots,0]P^{-1}, \tag{2}$$

式中，1对应于$\varphi_k(\lambda_k)$的位置，0对应于$\varphi_k(\lambda_j)(\lambda_j\neq\lambda_k)$的位置. 由此也容易推出$\varphi_1(A),\cdots,\varphi_s(A)$的线性无关性，并且，

$$\varphi_1(A)+\varphi_2(A)+\cdots+\varphi_s(A)=I_n,$$

$$(\varphi_k(A))^2=\varphi_k(A),\varphi_k(A)\varphi_j(A)=O\qquad(k\neq j).$$

基于上式中第一式，我们称$\varphi_1(A),\cdots,\varphi_s(A)$为关于简单矩阵$A$的**单位的分解**；余下两式表明，所有的$\varphi_k(A)$为彼此正交的幂等矩阵（投影）. 特殊地，若$A\in\boldsymbol{M}_n(\mathbb{C})$为正规的，(2)式中$P$可取为$U\in\mathscr{U}_n$，那么这时每一个$\varphi_k(A)$还为Hermite阵，因而它们均为正交投影（见第一章1.2.3小节）.

取 $E_k=\varphi_k(A)$，$k=1,\cdots,s$. 综上所述我们得到下面定理的第一部分与最后部分的结论.

定理 1($f(A)$的谱分解定理) 设简单矩阵 $A\in \boldsymbol{M}_n(\mathbb{C})$有 s 个不同特征值 $\lambda_1,\cdots,\lambda_s$，且 $f\in\mathscr{H}$. 则 $f(A)$为 $\boldsymbol{M}_n(\mathbb{C})$内的简单矩阵，且存在与 f 无关的 $E_k\in \boldsymbol{M}_n(\mathbb{C})$，$k=1,\cdots,s$，使得

$$E_1+E_2+\cdots+E_s=I_n,\quad E_k^2=E_k, E_kE_j=O(k\neq j);\tag{3}$$

$$f(A)=\sum_{j=1}^{s}f(\lambda_j)E_j.\tag{4}$$

并且当诸 $f(\lambda_j)$互不相等时，这些 E_k 是唯一的：$E_k\equiv\varphi_k(A)$，$k=1,\cdots,s$. 特殊地，若 A 为正规的，则 $f(A)$也是正规的，且 E_k 是正交投影，$k=1,\cdots,s$.

证明 余下只要证明(4)式中 $E_1,\cdots,E_s$ 当诸 $f(\lambda_j)$彼此不等时的唯一性. 若 $F_1,\cdots,F_s\in \boldsymbol{M}_n(\mathbb{C})$也满足条件(3)与(4)式：$f(A)=\sum\limits_{j=1}^{s}f(\lambda_j)F_j$，则有 $F_jf(A)=f(A)F_j=f(\lambda_j)F_j$，$\forall j$，由此并应用 $E_if(A)=f(A)E_i=f(\lambda_i)E_i\,(i=1,\cdots,s)$的事实，我们推出，$E_if(A)F_j=f(\lambda_j)E_iF_j=f(\lambda_i)E_iF_j$. 但 $f(\lambda_i)\neq f(\lambda_j)$，$\forall i\neq j$，于是，$E_iF_j=O$，$\forall i\neq j$. 因此，$E_i=E_i\sum\limits_{j=1}^{s}F_j=E_iF_i=\sum\limits_{j=1}^{s}E_jF_i=F_i$，$\forall i$. □

若取 $f(\lambda)=\lambda\in\mathscr{H}$，定理 1 便退化为简单矩阵 A 或正规矩阵 A 的谱定理(见第一章 1.1.1 小节定理 7 或 1.2.1 小节定理 5)的改进形式：

$$A=\sum_{j=1}^{s}\lambda_jE_j.\tag{5}$$

这里，满足条件(3)与(5)式的 $E_1,\cdots,E_s$ 由 A 唯一确定.

通常，表达式(4)称为简单矩阵 $f(A)$的**谱形式**(spectral form). 由于 $f(\lambda_1)$，$\cdots$，$f(\lambda_s)$之中可能有相同的数，故严格地说，(4)式不一定是 $f(A)$的谱分解. 但只要将相同的 $f(\lambda_j)$合并在一起，便能得到 $f(A)$的谱分解. 例如，若诸 $f(\lambda_j)$中只有 $f(\lambda_1)=f(\lambda_2)$，其余彼此不等，则 $f(A)=f(\lambda_1)(E_1+E_2)+f(\lambda_3)E_3+\cdots+f(\lambda_s)E_s$ 按定义便是 $f(A)$的谱分解.

作为定理 1 的推论，我们有如下有用的结果.

推论 2 设 $A\in \boldsymbol{M}_n(\mathbb{C})$为正规矩阵，且 $f,\bar{f}\in\mathscr{H}$，这里 $\bar{f}(\lambda)\equiv\overline{f(\lambda_j)}$. 则

$$\bar{f}(A)=f(A)^*,\tag{6}$$

$$\|f(A)\|_2=\max\{|f(\lambda_j)|:\lambda_j\in\sigma(A)\}.\tag{7}$$

证明 按(4)式，当 A 为正规矩阵时，$f(A)^*=\sum\limits_{j=1}^{s}\overline{f(\lambda_j)}E_j^*=\sum\limits_{j=1}^{s}\bar{f}(\lambda_j)E_j$. 因此，$\bar{f}(A)=f(A)^*$. 这时，$f(A)$有形如 3.1.1 小节(2)式的表示，但其中 P 为某个 $U\in\mathscr{U}_n$. 因此，

$$\begin{aligned}\| f(A) \|_2^2 &= \rho(f(A)^* f(A)) \\ &= \rho(U\mathrm{diag}[|f(\lambda_1)|^2,\cdots,|f(\lambda_n)|^2]U^*) \\ &= \max_j\{|f(\lambda_j)|^2\}.\end{aligned}$$ □

命题 3　应用定理 1 证明 Cayley 变换定理(第一章 1.2.1 小节定理 11).

证明　设 Hermite 矩阵 $A\in \boldsymbol{M}_n(\mathbb{C})$ 的谱为 $\sigma(A)=\{\lambda_1,\cdots,\lambda_s\}\subset\mathbb{R}$,且 A 有谱分解(5).此时,函数 $f(\lambda)=(\lambda-\mathrm{i})/(\lambda+\mathrm{i})\in\mathscr{H}$,且 f 将$\mathbb{R}$ 映射入复平面的单位圆周但 $f(\lambda)\neq 1$. 因此,$f(A)=(A-\mathrm{i}I_n)(A+\mathrm{i}I_n)^{-1}$ 有定义,且有谱形式 $f(A)=\sum_{j=1}^{s} f(\lambda_j)E_j$,其中 $|f(\lambda_j)|=1$ 但 $f(\lambda_j)\neq 1, j=1,\cdots,s$. 于是,$f(A)^* f(A)=f(A)f(A)*=\sum_{j=1}^{s}|f(\lambda_j)|^2E_j=I_n$,即 $f(A)=(A-\mathrm{i}I_n)(A+\mathrm{i}I_n)^{-1}\in\mathscr{U}_n$,且 $1\notin\sigma(f(A))$. 反之,设 $U\in\mathscr{U}_n, 1\notin\sigma(U)$,且其谱分解为 $U=\sum_{j=1}^{p}\xi_jF_j$. 考虑 $g(\lambda)=(\lambda+1)/(\mathrm{i}(\lambda-1))$. 显然,$g$ 在 $\sigma(U)$ 上有定义,且 g 将复平面单位圆周 $|\lambda|=1$(除去 $\lambda=1$) 映射入$\mathbb{R}$. 于是,$g(U)=\frac{1}{\mathrm{i}}(U+I_n)(U-I_n)^{-1}$ 有定义,且按定理 1,$g(U)$ 有谱形式

$$g(U)=\sum_{j=1}^{p}g(\xi_j)F_j,$$

式中 $g(\xi_j)\in\mathbb{R}, j=1,\cdots,p$. 因此,$g(U)^*=g(U)$,即 $g(U)=\frac{1}{\mathrm{i}}(U+I_n)(U-I_n)^{-1}$ 为 $\boldsymbol{M}_n(\mathbb{C})$ 中 Hermite 矩阵. □

仿照命题 3 的证明技巧,容易得到另一个基本结果.

命题 4　若 $A\in\boldsymbol{M}_n(\mathbb{C})$ 为 Hermite 矩阵,则 $U=\mathrm{e}^{\mathrm{i}A}\in\mathscr{U}_n$. 反之,任一 $U\in\mathscr{U}_n$ 有形式 $U=\mathrm{e}^{\mathrm{i}A}$,式中 A 可取为 Hermite 矩阵.

为证此命题,只需考虑复值函数 $f(\lambda)=\mathrm{e}^{\mathrm{i}\lambda}$ 即可(见习题 2). 显然,这是如下事实的推广:$|\mathrm{e}^{\mathrm{i}\alpha}|=1, \forall\alpha\in\mathbb{R}$.

下一命题为第一章 1.2.2 小节定理 4 的推广.

命题 5　若 $A\in\boldsymbol{M}_n(\mathbb{C})$ 为 Hermite 正半定(正定)矩阵,且 $k\geqslant 2$ 为给定的整数,则存在唯一的 Hermite 正半定(正定)矩阵 B 使得 $B^k=A$,且 $BA=AB$.

证明　考虑复值函数 $f(\lambda)=\lambda^{1/k}$ 的主单值分枝. 显然,$f\in\mathscr{H}$. 现设 A 有谱分解(5),式中,各 $\lambda_j\geqslant 0(>0)$. 这时,$B\equiv f(A)=\sum_{j=1}^{s}\lambda_j^{1/k}E_j$ 即为与 A 可交换的 A 的 Hermite 正半定(正定)k 次方根. 仿照第一章 1.2.2 小节定理 4 的证明,我们也可得到这样矩阵 B 的唯一性结论. □

通常将 $A\geqslant O$ 的唯一 Hermite 正半定 k 次方根记为 $A^{1/k}$.

命题 6　若 $A,B\in \boldsymbol{M}_n(\mathbb{C})$ 为 Hermite 正半定的，且 A 与 B 可交换，则 $AB=BA$ 也是 Hermite 正半定的.

证明　应用命题 5，$AB=BA$ 蕴涵 $A,B,A^{1/2},B^{1/2}$ 的两两可换性，于是，$AB=A^{1/2}B^{1/2}A^{1/2}B^{1/2}=(A^{1/2}B^{1/2})^2$. 但 $A^{1/2}$ 与 $B^{1/2}$ 为 Hermite 矩阵，且彼此可换，因而 $A^{1/2}B^{1/2}$ 也是 Hermite 矩阵. 由于 Hermite 矩阵的平方是正半定的，故 $AB=BA$ 为 Hermite 正半定的. □

除了命题 6 以外，定理 1 还可用以讨论涉及矩阵可换性的另外一些问题. 下面命题 7 与命题 8 便是这些问题中的重要例子.

命题 7　设 $A\in \boldsymbol{M}_n(\mathbb{C})$ 为正规的，且 $B\in \boldsymbol{M}_n(\mathbb{C})$ 满足 $AB=BA$. 则 $A^*B=BA^*$.

证明　设 A 有谱分解(5)，其中各 E_j 为 Hermite 矩阵，则 $A^*=\sum\limits_{j=1}^{s}\bar{\lambda}_jE_j$. 现令 $p\in \mathscr{P}_{s-1}$ 满足 $p(\lambda_j)=\bar{\lambda}_j,j=1,\cdots,s$. 于是，$A^*=p(A)$，即 A^* 可表示为 A 的多项式. 此事实连同 A 与 B 可换性假定立即推出 $A^*B=BA^*$. □

实际上可以类似地证明更深入的结果：若 $A,C\in \boldsymbol{M}_n(\mathbb{C})$ 为正规的，且 $B\in \boldsymbol{M}_n(\mathbb{C})$ 满足 $AB=BC$，则 $A^*B=BC^*$. 事实上，仿照命题 7 推证，有复多项式 q 使得 $A^*=q(A)$ 与 $C^*=q(C)$. 同时，$AB=BC$ 蕴涵 $A^kB=BC^k,k=1,2,\cdots$，因而 $q(A)B=Bq(C)$，即 $A^*B=BC^*$.

我们顺便地指出，命题 7 及其后面的推广都可视为复 Hilbert 空间算子理论中关于正规算子著名的 Fuglede—Putnam 定理的有限维情形(见文献[2]).

命题 8　设 $A,B\in \boldsymbol{M}_n(\mathbb{C})$ 为 Hermite 矩阵. 则 $AB=BA$ 当且仅当有 Hermite 矩阵 $C\in \boldsymbol{M}_n(\mathbb{C})$ 与两个实值函数 f 与 g，使得 $A=f(C)$ 与 $B=g(C)$. 并且，假如这样矩阵 C 存在，那么 C 可取为 $C=h(A,B)$，这里 h 为两个实变量的实系数多项式.

证明　条件充分性是显然的，因为按 3.1.1 小节定义 1，$f(C)g(C)=g(C)f(C)$，所以 $AB=BA$. 反之，设(5)式与 $B=\sum\limits_{i=1}^{p}\xi_iF_i$ 分别为 A 与 B 的谱分解，其中各 $\lambda_j,\xi_i\in\mathbb{R}$. 由于 $AB=BA$，故有 $E_jF_i=F_iE_j,j=1,\cdots,s;i=1,\cdots,p$. 现设 $h(r,t)=\sum\limits_k\sum\limits_l\alpha_{kl}r^kt^l$ 为实变量 r 与 t 的实系数多项式使得 $h(\lambda_j,\xi_i)=\zeta_{ji}$ 彼此不同，$j=1,\cdots,s;i=1,\cdots,p$. 记 $C=h(A,B)=\sum\limits_k\sum\limits_l\alpha_{kl}A^kB^l$. 则 $C=\sum\limits_j\sum\limits_i\left(\sum\limits_k\sum\limits_l\alpha_{kl}\lambda_j^k\xi_i^l\right)E_jF_i=\sum\limits_j\sum\limits_i h(\lambda_j,\xi_i)E_jF_i$. 这里 sp 个 E_jF_i 满足 $\sum\limits_{j,i}E_jF_i=I_n,(E_jF_i)^2=E_jF_i,(E_jF_i)(E_kF_l)=0$，只要 $(j,i)\neq(k,l)$，与 $(E_jF_i)^*=E_jF_i$. 因此，$C=C^*$. 选取 f 与 g 分别为满足 $f(\zeta_{ji})=\lambda_j$ 与 $g(\zeta_{ji})=\xi_i,j=1,\cdots,s;i=1,\cdots,p$ 的单变量实系数多项式. 这时，$f(C)=\sum\limits_i\sum\limits_i f(\zeta_{ji})E_jF_i=\sum\limits_j\sum\limits_i\lambda_jE_jF_i=$

$\sum_j \lambda_j E_j = A$，同理有 $g(C) = B$. □

习题 3.1.2

1. 求 3.1.1 小节例 4 中矩阵 A 的分量. 验证：它们满足(3)式，且为正交投影.

2. 证明命题 4.

3. 详细地证明命题 7 后面的推广结果.

4. 设 $A=UP$ 为 $A\in \boldsymbol{M}_n(\mathbb{C})$ 的极分解，其中 $U\in \mathcal{U}_n$，P 为 Hermite 正半定矩阵. 试证：A 正规当且仅当 $UP=PU$.（注意，由于 U 不必由 A 唯一确定，故本命题应理解为：若 A 正规，则 P 与每一个这样的 U 可换，反之，若 P 与某个这样的 U 可换，则 A 正规）

5. 设 $A\in \boldsymbol{M}_n(\mathbb{C})$ 的对角元素都为 a，所有非对角元素为 1. 试证：A 有单重特征值 $\lambda_1=n+a-1$ 与 $n-1$ 重特征值 $\lambda_2=a-1$. 试求 A 的分量.

6. 设 E_j，$j=1,\cdots,s$，为简单矩阵 $A\in \boldsymbol{M}_n(\mathbb{C})$ 的分量. 试证：$\mathrm{Im}E_j=\mathrm{Ker}(A-\lambda_j I)$，$\forall \lambda_j\in \sigma(A)$.

3.2　一般矩阵的函数

在 3.1 节中应用插值多项式技巧将简单矩阵 A 的函数 $f(A)$ 归结为该矩阵的某个多项式函数 $p(A)$. 在本节对一般矩阵的情形，仍沿用这个技巧给出一般矩阵函数 $f(A)$ 的多项式函数的等价形式，并讨论 $f(A)$ 的有关性质.

3.2.1　一般定义与性质

设 $A\in \boldsymbol{M}_n(\mathbb{C})$ 有表达式：

$$A = SJS^{-1}, \qquad \det S \neq 0, \tag{1}$$

式中，$J=\mathrm{diag}[J_1,\cdots,J_l]$ 为 A 的 Jordan 正规形式，诸 J_j 为 Jordan 块. 由第一章习题 1.1.2 题 9 知道，若 J_j 为对应于 A 的特征值 λ_i 的 r 阶 Jordan 块，则对任意复系数多项式 $p(\lambda)$，$p(J_j)$ 有如下的表达式：

$$p(J_j) = \begin{bmatrix} p(\lambda_i) & p'(\lambda_i) & \cdots & p^{(r-1)}(\lambda_i)/(r-1)! \\ 0 & p(\lambda_i) & \cdots & p^{(r-2)}(\lambda_i)/(r-2)! \\ \vdots & \ddots & \ddots \quad \ddots & \vdots \\ \vdots & & & p'(\lambda_i) \\ 0 & \cdots & 0 & p(\lambda_i) \end{bmatrix} \in \boldsymbol{M}_r(\mathbb{C}), \tag{2}$$

式中 $p^{(t)}(\lambda_i)$ 表示 $p(\lambda)$ 在 $\lambda=\lambda_i$ 处的 t 阶导数值（$p^{(0)}(\lambda)=p(\lambda)$）. 因此，$p(A)=p(SJS^{-1})=Sp(J)S^{-1}=S\mathrm{diag}[p(J_1),\cdots,p(J_l)]S^{-1}$. 这表明，$p(A)$ 与 A 的 Jordan 正规形式结构以及复多项式 $p(\lambda)$ 在 A 的特征值处函数值与各阶导数值有关，并且，这些导数的最高阶数恰等于 A 的最小多项式中对应的特征值的指标（即 A

的 Jordan 正规形式中对应于该特征值的最大 Jordan 块的阶数)减 1. 一般地说,若给定 $A\in \boldsymbol{M}_n(\mathbb{C})$,且知道它的最小多项式为

$$m(\lambda) = (\lambda-\lambda_1)^{m_1}(\lambda-\lambda_2)^{m_2}\cdots(\lambda-\lambda_s)^{m_s}, \tag{3}$$

式中,$\lambda_1,\cdots,\lambda_s$ 为 A 的不同特征值,m_k 为特征值 λ_k 的指标,那么,对任意复值函数 $f(\lambda)$,只要存在 $f(\lambda_k),f'(\lambda_k),\cdots,f^{(m_k-1)}(\lambda_k),k=1,2,\cdots,s$,我们便说,复值函数 $\boldsymbol{f}$ **在 $\boldsymbol{\sigma(A)}$ 上有定义**,并称 $f(\lambda_k),f'(\lambda_k),\cdots,f^{(m_k-1)}(\lambda_k),k=1,2,\cdots,s$ 为 **$\boldsymbol{f}$ 在 $\boldsymbol{\sigma(A)}$ 上的值**. 特殊地,若 $A\in \boldsymbol{M}_n(\mathbb{C})$ 为简单矩阵,则有 $m_1=m_2=\cdots=m_s=1$,因而这里的两个专门术语为 3.1.1 小节的对应术语的推广.

受多项式情形的启发,我们可给出 $f(A)$ 的定义.

定义 1　设 $A\in \boldsymbol{M}_n(\mathbb{C})$ 有 Jordan 正规形式表示(1),且复值函数 f 在 $\sigma(A)$ 上有定义. 则定义

$$f(A) = Sf(J)S^{-1} = S\mathrm{diag}[f(J_1),\cdots,f(J_l)]S^{-1}, \tag{4}$$

式中, $f(J_j)$ 等于(2)式中的矩阵,但其中的 p 全换为 f,即

$$f(J_j) = \begin{bmatrix} f(\lambda_i) & f'(\lambda_i) & \cdots & f^{(r-1)}(\lambda_i)/(r-1)! \\ 0 & f(\lambda_i) & \cdots & f^{(r-2)}(\lambda_i)/(r-2)! \\ \vdots & \ddots & \ddots \quad \ddots & \vdots \\ \vdots & & & f'(\lambda_i) \\ 0 & \cdots & 0 & f(\lambda_i) \end{bmatrix}. \tag{5}$$

我们指出,这样定义的 $f(A)$ 的合理性首先表现在当 $f(\lambda)$ 本身为复系数多项式时,这里的 $f(A)$ 与前面的矩阵多项式表示完全一致. 其次,当 $A\in \boldsymbol{M}_n(\mathbb{C})$ 为简单矩阵时,这里的 $f(A)$ 定义又与 3.1.1 小节定义 1 完全一致. 此外,如同简单矩阵的情形,$f(A)$ 的上述定义还提供用合适的多项式 $p(\lambda)$ 对应的 $p(A)$ 替代 $f(A)$ 的可能性. 事实上,按定义 1,$f(A)$ 只与 f 在 $\sigma(A)$ 上的值有关,(由后面看出,$f(A)$ 与(1)式中的 S 的选取无关)因而如能求出一个尽可能简单的函数,例如复系数多项式 $p(\lambda)$,使得它在 $\sigma(A)$ 上的值恰等于 f 在 $\sigma(A)$ 上的值,那么便有 $f(A)=p(A)$. 下面讨论这样多项式的存在性问题,它与所谓 **Hermite 多项式插值问题**紧密相关.

给定 s 个不同复数 $\lambda_1,\lambda_2,\cdots,\lambda_s$,$s$ 个正整数 $m_1,m_2,\cdots,m_s$,$m=\sum\limits_{k=1}^{s}m_k$ 与 m 个数:

$$f_{k,0},f_{k,1},\cdots,f_{k,m_k-1}, \qquad k=1,2,\cdots,s,$$

要寻求 $p(\lambda)\in \mathscr{P}_{m-1}$,使得

$$p(\lambda_k)=f_{k,0},p^{(1)}(\lambda_k)=f_{k,1},\cdots,p^{(m_k-1)}(\lambda_k)=f_{k,m_k-1}, \\ k=1,2,\cdots,s. \tag{6}$$

这就是著名的 **Hermite 多项式插值问题**. 我们将证明,$\mathscr{P}_{m-1}$ 内这样的多项式 $p(\lambda)$ 是唯一地存在的. 今后称之为满足条件(6)的 **Hermite 插值多项式**.

首先讨论在 $\mathscr{P}_{m-1}$ 内满足条件(6)的插值多项式的存在性与具体构造. 受 3.1.1 小节(3)式的启发,可以将满足条件(6)的多项式 $p\in\mathscr{P}_{m-1}$ 表示为 m 个较为简单的 $m-1$ 次多项式 $\varphi_{kj}(\lambda)(k=1,\cdots,s;j=0,1,\cdots,m_k-1)$ 的线性组合:

$$p(\lambda)=\sum_{k=1}^{s}\sum_{j=0}^{m_k-1}f_{k,j}\varphi_{kj}(\lambda), \tag{7}$$

式中,$\varphi_{kj}(\lambda)\in\mathscr{P}_{m-1}$ 满足

$$\varphi_{kj}^{(r)}(\lambda_k)=\delta_{jr},r=0,1,\cdots,m_k-1, \tag{8}$$

$$\varphi_{kj}^{(r)}(\lambda_i)=0,i\neq k,r=0,1,\cdots,m_i-1. \tag{9}$$

(注意,(8)与(9)两式等价于 $\varphi_{kj}^{(r)}(\lambda_i)=\delta_{ki}\delta_{rj},i=1,\cdots,s;r=0,1,\cdots,m_i-1$)这些 m 个多项式 $\varphi_{kj}(\lambda)\in\mathscr{P}_{m-1}$ 称为对应的 Hermite 多项式插值问题的**基函数**. 这样的 m 个基函数一旦存在,则容易直接验证,由(7)式确定的 $p\in\mathscr{P}_{m-1}$ 便是满足条件(6)的插值多项式. 事实上,由(8)与(9)两式看出,$p^{(r)}(\lambda_t)=\sum_{k=1}^{s}\sum_{j=0}^{m_k-1}f_{kj}\varphi_{kj}^{(r)}(\lambda_t)=\sum_{j=0}^{m_k-1}f_{tj}\varphi_{tj}^{(r)}(\lambda_t)=f_{t,r},\ t=1,\cdots,s;r=0,1,\cdots,m_t-1$. 另一方面,设

$$\sum_{k=1}^{s}\sum_{j=0}^{m_k-1}\beta_{kj}\varphi_{kj}(\lambda)=0,\qquad \beta_{kj}\in\mathbb{C}.$$

取 $\lambda=\lambda_t\in\sigma(A)$,代入上式则得 $\beta_{t0}=0$;代入

$$\sum_{k=1}^{s}\sum_{j=0}^{m_k-1}\beta_{kj}\varphi_{kj}^{(r)}(\lambda)=0,\qquad r=1,2,\cdots,m_t-1$$

则得 $\beta_{tr}=0,r=1,2,\cdots,m_t-1$. 于是,这些 m 个基函数线性无关,因而组成 m 维复线性空间 $\mathscr{P}_{m-1}$ 的基.

这样一来,Hermite 插值多项式 $p(\lambda)$ 的存在性问题等价于满足条件(8)与(9)的基函数 $\varphi_{kj}(\lambda)$ 的存在性问题. 现在证明这些基函数的存在性. 按条件(9),对给定 k 与 j,$\lambda_i(i\neq k)$ 是 $\varphi_{kj}(\lambda)$ 的 m_i 重根,因而 $\varphi_{kj}(\lambda)$ 有形式

$$\varphi_{kj}(\lambda)=\alpha_k(\lambda)\prod_{\substack{i=1\\ i\neq k}}^{s}(\lambda-\lambda_i)^{m_i} \tag{10}$$

(若 $s=1$,记 $\varphi_{kj}(\lambda)=\alpha_k(\lambda)$),这里 $\alpha_k(\lambda)\in\mathscr{P}_{m_k-1}$ 与 j 有关,从而它可表为含有 m_k 个(与 j 有关的)待定系数 $\alpha_{k,0},\alpha_{k,1},\cdots,\alpha_{k,m_k-1}$ 的多项式:

$$\alpha_k(\lambda)=\alpha_{k,0}+\alpha_{k,1}(\lambda-\lambda_k)+\cdots+\alpha_{k,m_k-1}(\lambda-\lambda_k)^{m_k-1}. \tag{11}$$

取 $\lambda=\lambda_k$,则有

$$\varphi_{kj}(\lambda_k)=\delta_{j0}=\alpha_k(\lambda_k)\prod_{\substack{i=1\\ i\neq k}}^{s}(\lambda_k-\lambda_i)^{m_i}=\alpha_{k,0}\prod_{\substack{i=1\\ i\neq k}}^{s}(\lambda_k-\lambda_i)^{m_i}, \tag{12}$$

由此得出,

$$\alpha_{k,0} = \delta_{j0} \Big/ \prod_{i=1, i\neq k}^{s} (\lambda_k - \lambda_i)^{m_i}. \tag{13}$$

类似地，对 $1\leqslant r\leqslant m_k-1$，若记 $\psi_k(\lambda)=\prod\limits_{i=1,i\neq k}^{s}(\lambda-\lambda_i)^{m_i}$，则

$$\varphi_{kj}^{(r)}(\lambda_k) = \delta_{jr} = \sum_{t=0}^{r}\binom{r}{t}\alpha_k^{(t)}(\lambda_k)\psi_k^{(r-t)}(\lambda_k) = \sum_{t=0}^{r}\binom{r}{t}t\,!\,\alpha_{k,t}\psi_k^{(r-t)}(\lambda_k). \tag{14}$$

于是，在上式中取 $r=1$，因 $\psi_k(\lambda_k)\neq 0$，所以可求得 $\alpha_{k,1}$. 接着，取 $r=2$，便可由 $\alpha_{k,0}$ 与 $\alpha_{k,1}$ 求出 $\alpha_{k,2}$，如此下去便能求出(11)式中 m_k 个待定系数. 因此，我们已经具体地求出满足条件(8)与(9)的 $\varphi_{kj}(\lambda)$，$k=1,\cdots,s;j=0,1,\cdots,m_k-1$.

其次证明：在 $\mathscr{P}_{m-1}$ 中满足条件(6)的 Hermite 插值多项式是唯一的. 事实上，若 $q(\lambda)\in\mathscr{P}_{m-1}$ 也是这样的多项式，那么，$h(\lambda)=p(\lambda)-q(\lambda)\in\mathscr{P}_{m-1}$ 满足条件：$h(\lambda_k)=h^{(1)}(\lambda_k)=\cdots=h^{(m_k-1)}(\lambda_k)=0,k=1,2,\cdots,s$. 这表明 $h(\lambda)$ 有 $m=\sum\limits_{k=1}^{s}m_k$ 个根(计入重根)，因而 $h(\lambda)\equiv 0$.

特殊地，如 $m_1=m_2=\cdots=m_s=1$，那么条件(6)退化为 $p(\lambda_k)=f_{k0}$，$k=1,2,\cdots,s$，于是，对应的 Hermite 插值多项式(7)便退化为 3.1.1 小节中插值多项式(3)，而 $\varphi_{k0}(\lambda)$ 退化为那里的 $\varphi_k(\lambda)$，$k=1,2,\cdots,s$.

例 2　若 $A\in\boldsymbol{M}_n(\mathbb{C})$ 的最小多项式 $m(\lambda)$ 表达式(3)中，$m_1=m_2=\cdots m_s=2$，则条件(6)变为：$p(\lambda_k)=f_{k0}$，$p^{(1)}(\lambda_k)=f_{k1}$，$k=1,2,\cdots,s$. 此时，对应的 Hermite 插值多项式 $p(\lambda)$ 有形式：

$$p(\lambda) = \sum_{k=1}^{s}[f_{k0}\varphi_{k0}(\lambda) + f_{k1}\varphi_{k1}(\lambda)].$$

经计算容易求得(见(10)式～(14)式)

$$\varphi_{k0}(\lambda) = [1 - 2\varphi'_k(\lambda_k)(\lambda - \lambda_k)][\varphi_k(\lambda)]^2,$$
$$\varphi_{k1}(\lambda) = (\lambda - \lambda_k)[\varphi_k(\lambda)]^2, k = 1,2,\cdots,s,$$

式中 $\varphi_k(\lambda)$ 由 3.1.1 小节(4)式确定. □

若令 $f_{k0}=f(\lambda_k)$，$f_{k1}=f^{(1)}(\lambda_k)$，$\cdots$，$f_{k,m_{k-1}}(\lambda_k)=f^{(m_k-1)}(\lambda_k)$，则满足条件(6)的 Hermite 插值多项式 $p(\lambda)\in\mathscr{P}_{m-1}$ 与 $f(\lambda)$ 在 $\sigma(A)$ 上的值相同. 按前面分析，这时 $f(A)=p(A)$，确切地有

定理 3　设 $A\in\boldsymbol{M}_n(\mathbb{C})$ 有最小多项式(3)，且 f 在 $\sigma(A)$ 上有定义. 则有 $p(\lambda)\in\mathscr{P}_{m-1}$ 使得 $p^{(r)}(\lambda_k)=f^{(r)}(\lambda_k)$，$r=0,1,\cdots,m_k-1;k=1,2,\cdots,s$，并且

$$p(A) = f(A). \tag{15}$$

$f(A)$ 只与 f 在 $\sigma(A)$ 上的值有关，而与(1)式中矩阵 S 的选取无关.

我们指出，定理 3 显然是 3.1.1 小节定理 3 的推广，它表明可以用 A 的多项式函数 $p(A)$ 来确定 $f(A)$. 虽然这样可选取的多项式 $p(\lambda)$(不必限制在 $\mathscr{P}_{m-1}$ 内)

有无穷多个,但 $f(A)$ 为 $\boldsymbol{M}_n(\mathbb{C})$ 内唯一确定的矩阵.

例 4 设

$$A=\frac{\pi}{2}\begin{bmatrix}2&0&0\\0&1&1\\0&0&1\end{bmatrix},$$

求 $\sin A$.

令 $f(\lambda)=\sin\lambda$. A 的最小多项式 $m(\lambda)$ 为

$$m(\lambda)=(\lambda-\pi)(\lambda-\pi/2)^2.$$

因此,$\lambda_1=\pi, m_1=1, \lambda_2=\pi/2, m_2=2$. f 在 $\sigma(A)$ 上的值为 $f_{10}=f(\pi)=0, f_{20}=f(\pi/2)=1, f_{21}=f'(\pi/2)=0$. 这时,对应的 $\mathscr{P}_2$ 中 Hermite 插值多项式 $p(\lambda)=\frac{4}{\pi}\left(\lambda-\frac{1}{\pi}\lambda^2\right)$. 因此,

$$f(A)=\sin A=\frac{4}{\pi}\left(A-\frac{1}{\pi}A^2\right)=\begin{bmatrix}0&0&0\\0&1&0\\0&0&1\end{bmatrix}.$$ □

类似于简单矩阵情形,对给定 $A\in\boldsymbol{M}_n(\mathbb{C})$,用 $\mathscr{H}$ 表示在 $\sigma(A)$ 上有定义的复值函数的集合. 若在 $\mathscr{H}$ 内引入通常的加法、数乘与乘法运算,则 $\mathscr{H}$ 组成一个可交换的复代数. 定义映射 φ:

$$\varphi(f)=f(A),\qquad \forall f\in\mathscr{H}. \tag{16}$$

显然,φ 为自 $\mathscr{H}$ 到矩阵代数 $\boldsymbol{M}_n(\mathbb{C})$ 内的一个**同态**. 它将 $\mathscr{H}$ 内函数 $f(\lambda)=\lambda^p(p=0,1,\cdots)$ 映射到 $A^p(p=0,1,\cdots)$. 若 A 非奇异,则 φ 将 $f(\lambda)=\lambda^{-1}$ 映射到 A^{-1}. 事实上,这时 $f\in\mathscr{H}, \lambda f(\lambda)=1$. 若 $p(\lambda)$ 为对应于 $f(\lambda)$ 的满足条件(6)的 Hermite 插值多项式,则 $\lambda p(\lambda)$ 与 $\lambda f(\lambda)=1$ 在 $\sigma(A)$ 上有相等的值,因而 $Ap(A)=Af(A)=I_n$. 这表明 $f(A)=A^{-1}$.

现在讨论一般矩阵 $A\in\boldsymbol{M}_n(\mathbb{C})$ 的函数 $f(A)$ 的一些基本性质. 在本章后面,我们总假定 A 有 Jordan 正规形式表示(1)或 A 有最小多项式(3),$m=\sum\limits_{k=1}^{s}m_k$.

定理 5 若 $A\in\boldsymbol{M}_n(\mathbb{C})$ 为分块对角矩阵:

$$A=\operatorname{diag}[A_1,A_2,\cdots,A_t],$$

且 $f\in\mathscr{H}$,则

$$f(A)=\operatorname{diag}[f(A_1),f(A_2),\cdots,f(A_t)]. \tag{17}$$

证明 首先看出,对任意多项式 p, $p(A)=\operatorname{diag}[p(A_1),p(A_2),\cdots,p(A_t)]$. 现设 $q\in\mathscr{P}_{m-1}$ 为 f 在 $\sigma(A)$ 上的值的 Hermite 插值多项式. 按定理 3, $f(A)=q(A)=\operatorname{diag}[q(A_1),q(A_2),\cdots,q(A_t)]$. 由于 $\sigma(A_i)\subset\sigma(A), i=1,2,\cdots,t$, 故 q 与 f 在 $\sigma(A_i)$ 上有定义,且有相同的值,因而 $f(A_i)=q(A_i), i=1,2,\cdots,t$. 于是,(17)式成

立. □

我们指出，如同定理 5 的证明所示，应用定理 3，取与 f 在 $\sigma(A)$ 上有相同值的多项式 p，借助矩阵多项式 $p(A)$ 诸多明显的性质往往容易导出 $f(A)$ 的某些性质. 这是一种今后常用的技巧.

定理 6 设 $A\in \boldsymbol{M}_n(\mathbb{C})$，$f_1,f_2,\cdots,f_t\in\mathscr{H}$，且 $p(u_1,u_2,\cdots,u_t)$ 为关于 $u_1,u_2,\cdots,u_t$ 的数值多项式. 若函数 $f(\lambda)=p(f_1(\lambda),f_2(\lambda),\cdots,f_t(\lambda))$ 在 $\sigma(A)$ 上的值均为零，则

$$f(A)=p(f_1(A),f_2(A),\cdots,f_t(A))=O. \tag{18}$$

证明 假定 $p_1,p_2,\cdots,p_t$ 分别为 $f_1,f_2,\cdots,f_t$ 在 $\sigma(A)$ 上的值的 Hermite 插值多项式. 按定理 3，$p_j(A)=f_j(A)$，$j=1,2,\cdots,t$. 记 $q(\lambda)=p(p_1(\lambda),p_2(\lambda),\cdots,p_t(\lambda))$. 则多项式 q 与 f 在 $\sigma(A)$ 上显然有相同值，因而按题设 q 在 $\sigma(A)$ 上的值也均为零，于是，$f(A)=q(A)=O$，且

$$f(A)=p(f_1(A),\cdots,f_t(A)).$$ □

应用定理 6，我们可将某些著名的数值函数恒等式推广到矩阵函数的情形.

例如，取 $f_1(\lambda)=\sin\lambda$，$f_2(\lambda)=\cos\lambda$ 与 $p(u_1,u_2)=u_1^2+u_2^2-1$. 由于 f_1 与 f_2 为整函数，它们为定义在任意 $A\in\boldsymbol{M}_n(\mathbb{C})$ 谱 $\sigma(A)$ 上的复值函数，且 $p(f_1(\lambda),f_2(\lambda))=\sin^2\lambda+\cos^2\lambda-1=0$，$\forall\lambda\in\mathbb{C}$，故从(18)式推出，

$$\sin^2A+\cos^2A=I_n,\qquad \forall A\in\boldsymbol{M}_n(\mathbb{C}).$$

同理，取 $f_1(\lambda)=e^{\lambda}$ 与 $f_2(\lambda)=e^{-\lambda}$，$p(u_1,u_2)=u_1u_2-1$，我们便得

$$e^Ae^{-A}=I_n,\qquad \forall A\in\boldsymbol{M}_n(\mathbb{C}).$$

因此，对任意 $A\in\boldsymbol{M}_n(\mathbb{C})$，$e^A$ 非奇异且 $(e^A)^{-1}=e^{-A}$，此外，按定义 1 有 $e^O=I_n$.

后面两个定理(定理 7 与定理 9)描述 A 与 $f(A)$ 谱之间的关系.

定理 7(谱映射定理) 设 $A\in\boldsymbol{M}_n(\mathbb{C})$ 有特征值 $\lambda_1,\lambda_2,\cdots,\lambda_n$，且 $f\in\mathscr{H}$. 则 $f(A)$ 的特征值为 $f(\lambda_1),f(\lambda_2),\cdots,f(\lambda_n)$，即有 $\sigma(f(A))=f(\sigma(A))$.

证明 这是定义 1 的简单推论.

应用谱映射定理，我们可以导出一些有趣的结论.

例 8 (1) $\det e^A=e^{\operatorname{tr}A}$，$\forall A\in\boldsymbol{M}_n(\mathbb{C})$.

事实上，由定理 7，若 $\lambda_1,\cdots,\lambda_n$ 为 A 的特征值，则 e^A 特征值为 $e^{\lambda_1},e^{\lambda_2},\cdots,e^{\lambda_n}$，因而 $\det e^A=e^{\lambda_1+\cdots+\lambda_n}=e^{\operatorname{tr}A}$.

(2) 设 $A\in\boldsymbol{M}_n(\mathbb{C})$ 的谱 $\sigma(A)\subset\Delta\equiv\{\lambda\in\mathbb{C}:|\lambda|<1\}$，令 $C=(A+I)^{-1}(A-I)$. 则 $\sigma(C)\subset\{\lambda\in\mathbb{C}:\operatorname{Re}\lambda<0\}$.

事实上，令 $f(\lambda)=(\lambda-1)(\lambda+1)^{-1}$，则 $f\in\mathscr{H}$. 因此，$f(A)=C$，且 $\sigma(f(A))=\sigma(C)=f(\sigma(A))=\{(\lambda-1)(\lambda+1)^{-1}:\lambda\in\sigma(A)\}$. 但最后的集合包含在 $\{\lambda\in\mathbb{C}:\operatorname{Re}\lambda<0\}$ 内. □

我们指出. 谱映射定理只是得到 $\sigma(f(A))=f(\sigma(A))$，但此并不保证 $f(A)$ 与

A 有类似的 Jordan 正规形式或最小多项式的结构. 例如,在前面例 4 中,三阶矩阵 A 有最小多项式 $m_A(\lambda)=(\lambda-\pi)(\lambda-\pi/2)^2$,而 $f(A)$ 有最小多项式 $m_{f(A)}(\lambda)=(\lambda-f(\pi))(\lambda-f(\pi/2))=\lambda(\lambda-1)$,它不等于$(\lambda-f(\pi))(\lambda-f(\pi/2))^2$. 这表明,当 $r>1$ 时,对应于 $\lambda I-A$ 的特定的初等因子$(\lambda-\lambda_i)^r$ 经过由 $f(\lambda)$确定的变换,$(\lambda-f(\lambda_i))^r$不一定为$\lambda I-f(A)$ 的初等因子. 我们来考察其中的原因. 设 $A\in \boldsymbol{M}_n(\mathbb{C})$对应于特征值 λ_i 的最大阶数的 Jordan 块 J_j 有形式:

$$J_j=\begin{bmatrix}\lambda_i & 1 & & & O\\ & \lambda_i & 1 & & \\ & & \ddots & \ddots & \\ & & & & 1\\ O & & & & \lambda_i\end{bmatrix}\in \boldsymbol{M}_r(\mathbb{C}).$$

对 $f\in\mathscr{H}$来说,$f(J_j)$取形式(5). 此形式与 $f(J_j)$的 Jordan 正规形式一般可能相差甚远. 但是,特殊地若 $f'(\lambda_j)\neq 0$,则由(5)式给出的 $f(J_j)$的初等因子恰为$(\lambda-f(\lambda_j))^r$(见第一章习题 1.1.2 题 7). 这时,$\lambda I-f(A)$对应于特征值 $f(\lambda_i)$的初等因子可从 $\lambda I-A$ 对应于λ_i 的初等因子$(\lambda-\lambda_i)^r$ 得到,即只要将 λ_i 换成$f(\lambda_i)$即可. 另一种特殊情形出现在 $f^{(1)}(\lambda_i)=f^{(2)}(\lambda_i)=\cdots=f^{(r-1)}(\lambda_i)=0$ 的时候,这时 $f(J_i)$变为 r 阶对角矩阵 $\mathrm{diag}[f(\lambda_i),\cdots,f(\lambda_i)]$,因而$\lambda I-A$ 的初等因子$(\lambda-\lambda_i)^r$ 在变换下被分裂为$\lambda I-f(A)$的具有形式$\lambda-f(\lambda_i)$的 r 个初等因子. 前面例 4 中$\lambda I-A$ 的初等因子$(\lambda-\pi/2)^2$ 变换的情形就属于此类. 一般地,我们有下列结果.

定理 9　设 $A\in \boldsymbol{M}_n(\mathbb{C})$的最小多项式 $m(\lambda)$由(3)式确定,且 $f\in\mathscr{H}$. 则 $\lambda I-A$ 与$\lambda I-f(A)$的初等因子之间有如下关系:

$\lambda I-A$ 的初等因子	$\lambda I-f(A)$的初等因子
(1)$\lambda-\lambda_k$	$\lambda-f(\lambda_k)$
(2)$(\lambda-\lambda_k)^r,r>1$	$\lambda-f(\lambda_k)$,r 个,如果 $f^{(j)}(\lambda_k)=0,j,=1,2,\cdots,r-1$
(3)$(\lambda-\lambda_k)^r,r>1$	$(\lambda-f(\lambda_k))^p$,$l-q$ 个;$(\lambda-f(\lambda_k))^{p+1}$,$q$ 个. 这里,l 为小于 r 且使得 $f^{(l)}(\lambda_k)\neq 0$ 的最小正整数;整数 p 与 q 满足条件 $r=lp+q(0\leqslant q\leqslant l-1,p>0)$

证明　情形(1)显然;情形(2)前面已证. 情形(3)为第一章习题 1.1.2 题 8 的直接推论. □

本小节最后一个定理是关于矩阵的复合函数的重要结果.

定理 10　设 $A\in \boldsymbol{M}_n(\mathbb{C})$. 如果 f 是在$\sigma(A)$上有定义的复值函数,$\mu_k=f(\lambda_k)$,$k=1,\cdots,s$,而 g 为复值函数使得 $g^{(m_k-1)}(\mu_k)$存在,$k=1,\cdots,s$,则 $h(\lambda)=g(f(\lambda))$. 为在 $\sigma(A)$上有定义的复值函数,且 $h(A)=g(f(A))$.

证明　由于 f 定义在 $\sigma(A)$ 上，且 $f(\lambda_k)=\mu_k, k=1,2,\cdots,s$，应用复合函数求导法则于 $h(\lambda)=g(f(\lambda))$，按假定我们有

$$\begin{aligned}h(\lambda_k) &= g(\mu_k)\\ h'(\lambda_k) &= g'(\mu_k)f'(\lambda_k)\\ &\vdots\\ h^{(m_k-1)}(\lambda_k) &= g^{(m_k-1)}(\mu_k)[f'(\lambda_k)]^{m_k-1}+\cdots+g'(\mu_k)f^{(m_k-1)}(\lambda_k).\end{aligned}\tag{19}$$

由上式知道，h 也是定义在 $\sigma(A)$ 上的，且它在 $\sigma(A)$ 上的值由(19)式确定. 现设 $r(\mu)$ 为满足下列条件的某个多项式：

$$r^{(i)}(\mu_k)=g^{(i)}(\mu_k),\qquad i=0,1,\cdots,m_k-1;k=1,2,\cdots,s,\tag{20}$$

则由(19)式看出，$h(\lambda)$ 与 $r(f(\lambda))$ 在 $\sigma(A)$ 上的值相同. 按定理 6 的证明我们有

$$h(A)=r(f(A)).$$

余下只需证明：$r(B)=g(B)$，这里 $B=f(A)$. 为此只需证明：$\{g(\mu_k),g'(\mu_k),\cdots,g^{(m_k-1)}(\mu_k)\}_{k=1}^{s}$ 包含 $g(\mu)$ 在 $\sigma(B)$ 上的值. 这实际上等价于 B 的每个特征值 μ_k 的指标不超过 m_k，亦即：$p(\mu)\equiv(\mu-\mu_1)^{m_1}(\mu-\mu_2)^{m_2}\cdots(\mu-\mu_s)^{m_s}$ 为 B 的零化多项式. 为证实后一事实，令 $q(\lambda)=p(f(\lambda))=(f(\lambda)-f(\lambda_1))^{m_1}\cdots(f(\lambda)-f(\lambda_s))^{m_s}$. 则 $q(\lambda)$ 有零点 λ_k，它至少为 m_k 重根，$k=1,2,\cdots,s$. 因此，$q(\lambda)$ 在 $\sigma(A)$ 上的值均为零. 根据定理 6(取 $t=1$)，$q(A)=p(B)=p(f(A))=O$，因而 $p(\mu)$ 确为 B 的零化多项式. □

应用定理 10 可以得到矩阵函数的另一些有价值的恒等式. 例如，取 $f(\lambda)=2\lambda, g(\lambda)=\sin\lambda$，则 $h(\lambda)=g(f(\lambda))=\sin2\lambda=2\sin\lambda\cos\lambda$，按照定理 10 有

$$\sin2A=2\sin A\cos A,\qquad \forall A\in \boldsymbol{M}_n(\mathbb{C}).$$

又如，设 $A\in\boldsymbol{M}_n(\mathbb{C})$ 非奇异，则 A 在 $M_n(\mathbb{C})$ 中有对数，即 $\ln A$ 有定义. 事实上，由于 $0\notin\sigma(A)$，故 $f(\lambda)=\ln\lambda$ 为定义在 $\sigma(A)$ 上的复值函数，因而 $f(A)$ 有定义. 另一方面，取 $g(\lambda)=e^{\lambda}$，它为整函数，且 $g(f(\lambda))=e^{f(\lambda)}=\lambda$ 为定义在 $\sigma(A)$ 上的复值函数. 按定理 10，我们有 $A=e^{f(A)}$，于是，$f(A)$ 可视为非奇异矩阵 A 的对数 $\ln A$.

习题 3.2.1

1. 设

$$A=\begin{bmatrix}6&2&2\\-2&2&0\\0&0&2\end{bmatrix}.$$

求 $\ln A$.

2. 试证：若 B 为 Hermite 正定矩阵，则存在唯一的 Hermite 矩阵 H 使得 $B=e^H$. 试给出非 Hermite 矩阵 A 使得 e^A 为 Hermite 正定的例子.

3. 试证：若 $A\in\boldsymbol{M}_n(\mathbb{C})$，$p(\lambda)$ 与 $q(\lambda)$ 为相对互素多项式，且有 $\det q(A)\neq0$，则有理函数 $r(\lambda)=p(\lambda)/q(\lambda)$ 满足

$$r(A)=p(A)[q(A)]^{-1}=[q(A)]^{-1}p(A).$$

4. 试证：$e^{iA}=\cos A+i\sin A,\ \forall A\in \boldsymbol{M}_n(\mathbb{C})$.

5. 试证：若 f 为定义在 $A\in \boldsymbol{M}_n(\mathbb{C})$ 谱 $\sigma(A)$ 上的复值函数，则 $f(A^{\mathrm{T}})=[f(A)]^{\mathrm{T}}$.

6. 设 f 为定义在 Hermite 矩阵 H 谱 $\sigma(H)$ 上的复值函数，它满足 $|f(\lambda)|=1,\ \forall\lambda\in\mathbb{R}$. 试证：$f(H)$ 为酉矩阵.

3.2.2　一般矩阵函数的谱分解

设 $A\in \boldsymbol{M}_n(\mathbb{C})$ 的最小多项式 $m(\lambda)$ 由 3.2.1 小节(3)式确定，$f\in\mathscr{H}$. 应用 3.2.1 小节定理 3 与(7)式，我们有

$$f(A)=p(A)=\sum_{k=1}^{s}\sum_{j=0}^{m_k-1}f_{kj}\varphi_{kj}(A). \tag{1}$$

由前面讨论知道，这些 $m=\sum\limits_{k=1}^{s}m_k$ 个线性无关的 $\varphi_{kj}(\lambda)$ 只与 A 的最小多项式 $m(\lambda)$ 结构有关，而与 f 无关. 令 $\varphi_{kj}(A)=E_{kj}$. 我们有如下基本结果.

定理 1($f(A)$的谱分解定理)　设 $A\in \boldsymbol{M}_n(\mathbb{C})$ 有最小多项式 $m(\lambda)=(\lambda-\lambda_1)^{m_1}(\lambda-\lambda_2)^{m_2}\cdots(\lambda-\lambda_s)^{m_s}$，且 $f\in\mathscr{H}$. 则存在与 $f(\lambda)$ 无关的矩阵 $E_{kj}\in \boldsymbol{M}_n(\mathbb{C})$，$k=1,2,\cdots,s$；$j=0,1,\cdots,m_k-1$，使得

$$f(A)=\sum_{k=1}^{s}\sum_{j=0}^{m_k-1}f_{kj}E_{kj}. \tag{2}$$

并且，这些矩阵 E_{kj} 线性无关，它们与 A 可交换且彼此之间可交换.

证明　表达式(2)为(1)式的变型. 由于 $\varphi_{kj}(A)=E_{kj}$ 为 A 的多项式，故它们与 A，以及它们彼此之间可交换. 余下证明：矩阵 $E_{kj}(k=1,\cdots,s;j=0,1,\cdots,m_k-1)$ 线性无关.

假定复数 $c_{kj}(1\leqslant k\leqslant s,0\leqslant j\leqslant m_k-1)$ 满足

$$\sum_{k=1}^{s}\sum_{j=0}^{m_k-1}c_{kj}E_{kj}=O.$$

令 $h(\lambda)=\sum\limits_{k,j}c_{kj}\varphi_{kj}(\lambda)$，则 $h(\lambda)\in\mathscr{P}_{m-1}$，$m=\sum\limits_{k=1}^{s}m_k$. 于是，$h(A)=\sum\limits_{k,j}c_{kj}E_{kj}=O$. 这表明 $h(\lambda)$ 为次数不超过 $m-1$ 的 A 的零化多项式或为零函数，因而 $h(\lambda)\equiv0$，否则 A 有次数小于最小多项式 $m(\lambda)$ 的零化多项式. 再由 $\varphi_{kj}(\lambda)$ 的线性无关性直接推得所有 $c_{kj}=0$. □

我们指出，当 $A\in \boldsymbol{M}_n(\mathbb{C})$ 为简单矩阵时，各 $m_k=1$，因而上述 $f(A)$ 谱分解定理这时便简化为 3.1.2 小节中定理 1.

类似于简单矩阵函数情形，我们将(2)式中矩阵 E_{ij} 叫作 $A\in \boldsymbol{M}_n(\mathbb{C})$ 的**分量**. 这些 A 的分量中，

$$E_{k0}=\varphi_{k0}(A)=\prod_{\substack{j=1\\j\neq k}}^{s}(A-\lambda_j I)^{m_j}\Big/\prod_{\substack{j=1\\j\neq k}}^{s}(\lambda_k-\lambda_j)^{m_j}+\cdots,\qquad k=1,\cdots,s,$$

式中“…”表示关于 A 的某个多项式(但当所有 $m_j=1$ 时,它为零),并且,

$$\sum_{k=1}^{s}E_{k0}=I_n. \tag{3}$$

这后一等式得自(2)式,只要令那里的 $f(\lambda)=1$ 即可. 通常称它为**关于 $A\in M_n(\mathbb{C})$ 的单位的分解**. 若令(2)式中 $f(\lambda)=\lambda$ 便得

$$A=\sum_{k=1}^{s}(\lambda_k E_{k0}+(1-\delta_{1,m_k})E_{k1}). \tag{4}$$

这是简单矩阵谱定理的推广(见 3. 1. 2 小节(5)式).

为了考察 A 的分量的进一步性质,我们研究这些分量的结构形式. 若 $A\in M_n(\mathbb{C})$有形式 $A=SJS^{-1}=S\mathrm{diag}[J_1,J_2,\cdots,J_l]S^{-1}$,式中 J 为 A 的 Jordan 正规形式,按 3. 2. 1 小节定义 1,

$$E_{kj}=\varphi_{kj}(A)=S\mathrm{diag}[\varphi_{kj}(J_1),\cdots,\varphi_{kj}(J_l)]S^{-1}, \tag{5}$$

其中,$\varphi_{kj}(J_i)$具有 3. 2. 1 小节中形式(5)(但其中 f 用 φ_{kj} 替代),$1\leqslant i\leqslant l,1\leqslant k\leqslant s$, $0\leqslant j\leqslant m_k-1$. 对于 $\varphi_{kj}(J_i)$,讨论两种可能情形.

情形Ⅰ　J_i 是对应于 A 的特征值 λ_k 的阶数为 n_i 的 Jordan 块. 此时,由于 $\varphi_{kj}^{(r)}(\lambda_k)=\delta_{jr}(r,j=0,1,\cdots,m_k-1)$,故当 $j\neq 0$ 时有

$$\varphi_{kj}(J_i)=\overset{\qquad\qquad j+1\,列}{\begin{bmatrix}0 & 0 & \cdots & \frac{1}{j!} & 0\\ & 0 & & & \ddots\\ & & & \ddots & \frac{1}{j!}\\ & & \ddots & & \\ & & & & \vdots\\ & & & & 0\\ & 0 & & & 0\end{bmatrix}}=\frac{1}{j!}N_{n_i}^{j}, \tag{6}$$

式中,N_{n_i} 为 n_i 阶移位幂零矩阵,它的$(i,i+1)$元素$(i=1,\cdots,n_i-1)$为 1,其余均为 0. 当 $j=0$ 时,$\varphi_{kj}(J_i)=I_{n_i}$. 按 λ_k 指标 m_k 的定义,显然有 $n_i\leqslant m_k$.

情形Ⅱ　J_i 是对应于 A 的特征值 $\lambda_l(\neq\lambda_k)$的阶数为 n_i 的 Jordan 块. 这时有 $n_i\leqslant m_l$,且 $\varphi_{kj}^{(r)}(\lambda_l)=0,r=0,1,\cdots,m_l-1$. 因此,$\varphi_{kj}(J_i)=O$.

将上述两种情形关于 $\varphi_{kj}(J_i)$的结果代入(5)式,便得

$$E_{kj}=\varphi_{kj}(A)=\frac{1}{j!}S\mathrm{diag}[O,\cdots,O,N_{n_1}^{j},N_{n_2}^{j},\cdots,N_{n_q}^{j},O,\cdots,O]S^{-1}, \tag{7}$$

这里我们不妨假定 $J_{n_1},J_{n_2},\cdots,J_{n_q}$ 为 A 对应于特征值 λ_k 的所有 Jordan 块. 特别

地，当取 $j=0$ 时，便有

$$E_{k0}=\varphi_{k0}(A)=S\operatorname{diag}[O,\cdots,O,I_r,O,\cdots,O]S^{-1}, \tag{8}$$

式中 $r=n_1+n_2+\cdots+n_q$ 恰为特征值 λ_k 的代数重数.

应用 A 的分量 E_{kj} 的明显表达式(7)与(8)，我们可以得到下面重要结果.

定理 2 矩阵 $A\in \boldsymbol{M}_n(\mathbb{C})$ 的分量 E_{kj} $(1\leqslant k\leqslant s, 0\leqslant j\leqslant m_k-1)$ 具有下列性质：

(1) $\sum\limits_{k=1}^{s}E_{k0}=I.$

(2) $\sum\limits_{k=1}^{s}(\lambda_kE_{k0}+(1-\delta_{1,m_k})E_{k1})=A.$

(3) 当且仅当 $j=0$ 时，$E_{kj}^2=E_{kj}$（即诸 E_{kj} 中只有 E_{k0} 为投影矩阵）.

(4) 如果 $k\neq l$，则 $E_{kp}E_{lq}=O.$

(5) $E_{k0}E_{kj}=E_{kj}, j=0,1,\cdots,m_k-1.$ 并且一般地，

$$E_{kl}E_{kj}=\begin{cases}\dfrac{(j+l)!}{j!l!}E_{k,l+j}, & \text{如果 } l+j\leqslant m_k-1,\\ O, & \text{如果 } l+j>m_k-1.\end{cases}$$

证明 应用 E_{kj} 表达式(7)与(8)直接计算便得(3)～(5). (1)与(2)即为(3)式与(4)式. □

我们指出，前一定理中(1)与(2)也可以由(7)与(8)两式推出. 通常为简单起见，约定当 $m_k=1$ 时，$E_{k1}=O$，于是(4)式或定理 2(2)可改写为

$$A=\sum_{k=1}^{s}(\lambda_kE_{k0}+E_{k1}). \tag{4'}$$

将上式两端右乘以 E_{k0}，并应用定理 2(3)～(5)容易导出，$E_{k1}=(A-\lambda_kI)E_{k0}$. 接着用 E_{k1} 右乘以最后等式便得，$E_{k2}=\dfrac{1}{2!}(A-\lambda_kI)E_{k1}=\dfrac{1}{2!}(A-\lambda_kI)^2E_{k0}$, $k=1,\cdots,s$. 类似地可归纳地证明如下更一般的结果.

推论 3 设 E_{kj} $(1\leqslant k\leqslant s, 0\leqslant j\leqslant m_k-1)$ 为 $A\in \boldsymbol{M}_n(\mathbb{C})$ 的分量. 则

$$E_{kj}=\frac{1}{j!}(A-\lambda_kI)^jE_{k0}. \tag{9}$$

表达式(8)另一个有价值的推论为下面结果.

推论 4 $\operatorname{Im}E_{k0}=\operatorname{Ker}(A-\lambda_kI)^{m_k}, k=1,2,\cdots,s.$

证明 从(8)式与第一章 1.2.3 小节定理 1(3)看出，由于 $S\in \boldsymbol{M}_n(\mathbb{C})$ 非奇异，$\operatorname{Im}E_{k0}=\operatorname{Ker}(I-E_{k0})=\operatorname{Ker}\operatorname{diag}[I,\cdots,I,O_r,I,\cdots,I]S^{-1}$, $r=n_1+\cdots+n_q$. 另一方面，$(A-\lambda_kI)^{m_k}=S\operatorname{diag}[*,\cdots,*,O_r,*,\cdots,*]S^{-1}$，式中各 $*$ 均为非奇异矩阵，O_r 对应于 λ_k 的 Jordan 块.（由于 $n_1,\cdots,n_q$ 均不超过 m_k，故 $(J_{n_t}-\lambda_kI_{n_t})^{m_k}=N_{n_t}^{m_k}=O, t=1,2,\cdots,q$）并且，$\operatorname{Im}E_{k0}$ 与 $(A-\lambda_kI)^{m_k}$ 的表达式中的两个块对角矩阵有相容的分块. 因此，$\operatorname{Ker}(A-\lambda_kI)^{m_k}$ 等于 $\operatorname{Ker}\operatorname{diag}[*,\cdots,*,O_r,*,\cdots,*]S^{-1}$. 于是，

$\mathrm{Im}E_{k0}=\mathrm{Ker}(A-\lambda_k I)^{m_k}, k=1,2,\cdots,s.$ □

我们指出，推论 4 直接蕴涵，$(A-\lambda_k I)^{m_k}E_{k0}\boldsymbol{x}=\boldsymbol{0}, \forall \boldsymbol{x}\in\mathbb{C}^n$，因而

$$(A-\lambda_k I)^{m_k}E_{k0}=O, \qquad k=1,2,\cdots,s. \tag{10}$$

但是，对于 A 的其他分量 $E_{kj}(j\neq 0)$，一般没有推论 4 的推广结果，按(9)式有如下结论：

$$\mathrm{Im}E_{k0}\supset \mathrm{Im}E_{k_1}\supset\cdots\supset \mathrm{Im}E_{k,m_k-1}, \tag{11}$$

并且，$\mathrm{Im}E_{k,m_k-1}$ 为 $\mathrm{Ker}(A-\lambda_k I)$ 的一个子空间，记为 $\mathscr{S}_k$，$\dim\mathscr{S}_k$ 等于 A 的 Jordan 正规形式中对应于特征值 λ_k 并有阶数 m_k 的 Jordan 块的数目（见习题 9）.

现在考虑定理 1 与定理 2 的应用例子.

例 5 设 $A\in\boldsymbol{M}_n(\mathbb{C})$. 则有

$$A^2=\sum_{k=1}^{s}(\lambda_k^2E_{k0}+2\lambda_kE_{k1}+2E_{k2}),$$

$$A^l=\sum_{k=1}^{s}(\lambda_k^lE_{k0}+l\lambda_k^{l-1}E_{k1}+l(l-1)\lambda_k^{l-2}E_{k2}+\cdots+l!E_{kl}), l=1,2,\cdots.$$

事实上，取 $f(\lambda)=\lambda^2$ 或 λ^l，应用定理 1 直接求得这两个表达式. 同样地，应用定理 2 中 A 分量之间的关系也容易计算出 A^2，再用数学归纳法便得 A^l 的表达式. □

定理 1 与定理 2 常有助于计算 $f(A)$，特别当 f 取 $\mathscr{H}$ 中不同复值函数时，按公式(2)仅需变换一下系数 f_{kj}，而不必重新计算 A 的分量 E_{kj}. 下面例 6 告诉我们，一般可以不必借助于 Hermite 插值问题的基函数 $\varphi_{kj}(\lambda)$ 来确定 A 的分量 $E_{kj}=\varphi_{kj}(A)$. 注意，此例中提供的求 A 分量的办法具有相当的普遍性.

例 6 设

$$A=\begin{bmatrix}3 & -1 & 1\\ 2 & 0 & 1\\ 1 & -1 & 2\end{bmatrix},$$

且 $\alpha\in\mathbb{C}$ 为常数，求 A 的分量与 $\mathrm{e}^{\alpha A}$ 以及 A^{100}.

A 的最小多项式为 $m(\lambda)=(\lambda-1)(\lambda-2)^2$. 取 $\lambda_1=1, m_1=1, \lambda_2=2, m_2=2, m=m_1+m_2=3$. 对任意 $f\in\mathscr{H}$，定理 1 给出，

$$f(A)=f(1)E_{10}+f(2)E_{20}+f^{(1)}(2)E_{21}. \tag{12}$$

为了确定出 E_{10}, E_{20}, E_{21}，选取三个线性无关的多项式分别代入(12)式便推导出关于 E_{10}, E_{20} 与 E_{21} 的方程组，求解之可得 E_{10}, E_{20} 与 E_{21} 的具体矩阵形式. 譬如，依次取 $f(\lambda)=1, f(\lambda)=\lambda-2$ 与 $f(\lambda)=(\lambda-2)^2$，代入(12)式，则得

$$E_{10}+E_{20}=I_3$$
$$-E_{10}+E_{21}=A-2I_3$$
$$E_{10}=(A-2I_3)^2.$$

求解方程组便有

$$E_{10}=\begin{bmatrix}0&0&0\\-1&1&0\\-1&1&0\end{bmatrix},\quad E_{20}=\begin{bmatrix}1&0&0\\1&0&0\\1&-1&1\end{bmatrix},\quad E_{21}=\begin{bmatrix}1&-1&1\\1&-1&1\\0&0&0\end{bmatrix}.$$

因此，

$$e^{\alpha A}=e^{\alpha}E_{10}+e^{2\alpha}E_{20}+\alpha e^{2\alpha}E_{21}=\begin{bmatrix}(\alpha+1)e^{2\alpha}&-\alpha e^{2\alpha}&\alpha e^{2\alpha}\\-e^{\alpha}+(\alpha+1)e^{2\alpha}&e^{\alpha}-\alpha e^{2\alpha}&\alpha e^{2\alpha}\\-e^{\alpha}+e^{2\alpha}&e^{\alpha}-e^{2\alpha}&e^{2\alpha}\end{bmatrix};$$

$$A^{100}=E_{10}+2^{100}E_{20}+100\times 2^{99}E_{21}=\begin{bmatrix}a+b&-b&b\\-1+a+b&1-b&b\\-1+a&1-a&a\end{bmatrix},$$

其中，$a=2^{100}$，$b=100\times 2^{99}$. □

回顾一下，当 $A\in\boldsymbol{M}_n(\mathbb{C})$ 为正规矩阵时，A 的所有分量 $E_1=E_{10},\cdots,E_s=E_{s0}$ 为 Hermite 幂等矩阵，即正交投影，且两两正交，即 $E_iE_j=O, i\neq j$. 确切地说，每个 $E_k=E_{k0}$ 为 $\mathrm{Ker}(A-\lambda_kI)=\mathrm{Im}E_k$ 上的正交投影，亦即 $E_k=E_{k0}$ 为对应于特征值 λ_k 特征空间 $\mathrm{Ker}(A-\lambda_kI)$ 上的正交投影. 并且，只要 $i\neq j$，便有 $\mathrm{Im}E_i$ 与 $\mathrm{Im}E_j$ 正交，因而 $C^n=\sum\limits_{j=1}^{s}\oplus\mathrm{Im}E_j=\sum\limits_{j=1}^{s}\oplus\mathrm{Ker}(A-\lambda_jI)$. 在 $A\in\boldsymbol{M}_n(\mathbb{C})$ 为一般矩阵情形，A 的分量 $E_{10},E_{20},\cdots,E_{s0}$ 虽然为幂等的（见定理 2(3)）但不一定为 Hermite 矩阵，因而不必是正交投影（如例 6 中的 E_{10} 与 E_{20}）. 下面定理表明，它们仍然有着较好的性质.

定理 7　设 $A\in\boldsymbol{M}_n(\mathbb{C})$ 有 s 个不同的特征值 $\lambda_1,\cdots,\lambda_s$，它们指标分别为 $m_1,\cdots,m_s$. 则对应于 λ_k 的 A 的分量 E_{k0} 是沿着不同于 λ_k 的特征值所对应的广义特征空间直和

$$\sum_{\substack{j=0\\ j\neq k}}^{s}\mathrm{Ker}(A-\lambda_jI)^{m_j}$$

在对应于 λ_k 的（m_k 级）广义特征空间 $\mathrm{Ker}(A-\lambda_kI)^{m_k}$ 上的投影，$k=1,\cdots,s$.

证明　由定理 2(3)，E_{k0} 为幂等矩阵，因而它是沿着 $\mathrm{Ker}E_{k0}$ 在 $\mathrm{Im}E_{k0}$ 上的投影，且 $\mathrm{Ker}E_{k0}=\mathrm{Im}(I-E_{k0})$（见(8)式或第一章 1.2.3 小节定理 1(3)）. 另一方面，由 $\mathrm{Im}E_{i0}\cap\mathrm{Im}E_{j0}=\{\boldsymbol{0}\}\ (i\neq j)$ 与 $\sum\limits_{j=1}^{s}E_{j0}=I$ 推出，$\mathrm{Ker}E_{k0}=\mathrm{Im}(I-E_{k0})=\mathrm{Im}\Big(\sum\limits_{j=1,j\neq k}^{s}E_{j0}\Big)=\sum\limits_{j=1,j\neq k}^{s}\dot{}\ \mathrm{Im}E_{j0}=\sum\limits_{j=1,j\neq k}^{s}\dot{}\ \mathrm{Ker}(A-\lambda_jI)^{m_j}$（见推论 4）. 再引用推论 4，$\mathrm{Im}E_{k0}=\mathrm{Ker}(A-\lambda_k)^{m_k}$. □

我们指出，应用公式(8)也能得出定理 7 的结论. 此外，从推论 4 看出，$\mathrm{Im}E_{k0}=$

$\operatorname{Ker}(A-\lambda_k I)^{m_k}$，因而 $\operatorname{Im}E_{j0}$ 与 $\operatorname{Ker}E_{k0}$ 皆为 A 的不变子空间，$k=1,2,\cdots,s$.

习题 3.2.2

1. 试证：若 $A\in \boldsymbol{M}_n(\mathbb{C})$ 与 $\lambda\notin\sigma(A)$，则

$$(\lambda I-A)^{-1}=\sum_{k=1}^{s}\sum_{j=0}^{m_k-1}\frac{j!}{(\lambda-\lambda_k)^{j+1}}E_{kj}.$$

2. 设 $A=SJS^{-1}$，其中 $J=\operatorname{diag}[J_1,J_2,\cdots,J_7]$，$J_1=-I_6+N_6$，$J_2=-I_6+N_6$，$J_3=-I_4+N_4$，$J_4=3I_3+N_3$，$J_5=3I_5+N_5$，$J_6=I_2+N_2$，$J_7=2I_4+N_4$，试求出 A 的所有分量 E_{kj}. 用求得的结果验证一下定理 2、推论 3、推论 4 与定理 7 的结论.

3. 设

$$A=\begin{bmatrix}6&2&2\\-2&2&0\\0&0&2\end{bmatrix}.$$

求 A 的分量，再用定理 1 求 $\ln A$.（试与习题 3.2.1 题 1 结果比较一下）

4. 设

$$A=\begin{bmatrix}-1&1&2&1\\0&-1&0&-3\\0&0&1&1\\0&0&0&0\end{bmatrix}.$$

求 A 的分量，再用定理 1 求 B 使得 $B^3=A$.

5. 设 $A\in\boldsymbol{M}_n(\mathbb{C})$ 非奇异或奇异但零特征值的指标为 1. 试证：对任意正整数 p，总存在 $B\in\boldsymbol{M}_n(\mathbb{C})$ 使得 $B^p=A$. 这样的 B 为唯一的吗？为什么？

6. 继续题 1，令

$$E_1=\lim_{\lambda\to\lambda_1}(\lambda I-A)^{-1}(I-E_{10}).$$

试证：$E_1=\sum_{k=2}^{s}\sum_{j=0}^{m_k-1}j!(\lambda_1-\lambda_k)^{-j-1}E_{kj}$ 且 $I-E_{10}=E_1(\lambda_1 I-A)$.

7. 设 $A\in\boldsymbol{M}_n(\mathbb{C})$ 有 s 个不同特征值 $\lambda_1,\lambda_2,\cdots,\lambda_s$. 又设 $f_k\in\mathscr{H}$ 使得 $f_k(\lambda)\equiv1$，$\forall\lambda\in C(\lambda_k,\varepsilon_k)$；$f_k(\lambda)\equiv0$，$\forall\lambda\in C(\lambda_j,\varepsilon_j)$，$j\neq k$，这里 $C(\lambda_i,\varepsilon_i)=\{\lambda\in\mathbb{C}:|\lambda-\lambda_i|\leqslant\varepsilon_i\}$，选取 $\varepsilon_i>0$ 足够小使得 s 个 $C(\lambda_i,\varepsilon_i)$ 彼此不相交. 试证：

$$f_k(A)=E_{k0},\qquad k=1,2,\cdots,s,$$

并由此推出 $E_{k0}^2=E_{k0}$，$k=1,2,\cdots,s$.

8. 设 $A\in\boldsymbol{M}_n(\mathbb{C})$ 同题 7. 试证：对任意 $\lambda\notin\sigma(A)$，

$$(\lambda I-A)^{-1}=\sum_{j=0}^{m_1-1}j!(\lambda-\lambda_1)^{-j-1}E_{1j}+E(\lambda),$$

式中 $E(\lambda)\in\boldsymbol{M}_n(\mathbb{C})$，它的元素在 λ_1 某邻域内皆解析，且有 $E_{1j}E(\lambda)=E(\lambda)E_{1j}=O$. 记 $E_1=E(\lambda_1)$，应用这个 $(\lambda I-A)^{-1}$ 的表达式推出前面题 6 中公式 $I-E_{10}=E_1(\lambda_1 I-A)$.

9. 在(11)式中令 $\mathscr{S}_k=\operatorname{Im}E_{k,m_k-1}$. 试证：$\dim\mathscr{S}_k$ 等于 A 的 Jordan 正规形式中对应于特征值 λ_k 并有阶数 m_k 的 Jordan 块的数目.

3.2.3　矩阵函数的序列与级数

在第二章 2.2.2 小节中,我们已经讨论了矩阵幂序列与级数的收敛性问题. 应用矩阵函数的概念可以研究一般的矩阵函数的序列与级数的收敛性问题. 确切地说,设 $A=(a_{i,j})\in \boldsymbol{M}_n(\mathbb{C})$, f_p 与 $u_p\in\mathscr{H}$, $p=1,2,\cdots$,要研究序列 $f_1(A),f_2(A),\cdots,f_p(A),\cdots$的收敛性与级数 $\sum\limits_{p=1}^{\infty}u_p(A)$ 的收敛性.

令 $A_p\equiv f_p(A)=(\alpha_{ij}^{(p)})\in\boldsymbol{M}_n(\mathbb{C})$, $p=1,2,\cdots$. 若将诸 A_p 看成 n^2 维线性空间 $\boldsymbol{M}_n(\mathbb{C})$中的向量,则按定义序列 $A_1,A_2,\cdots,A_p,\cdots$收敛于 $B=(b_{ij})\in\boldsymbol{M}_n(\mathbb{C})$当且仅当

$$\lim_{p\to\infty}\alpha_{ij}^{(p)}=b_{ij},\quad i,j=1,2,\cdots,n. \tag{1}$$

由于有限维向量空间上向量范数的等价性,(1)式等价于

$$\lim_{p\to\infty}\|A_p-B\|=0, \tag{2}$$

式中, $\|\cdot\|$ 可取为 $\boldsymbol{M}_n(\mathbb{C})$上任意的广义矩阵范数.

下面定理 1 表明,矩阵函数序列$\{f_p(A)\}_{p=1}^{\infty}$的收敛性等价于 $m=\sum\limits_{k=1}^{s}m_k$ 个数列$\{f_p^{(j)}(\lambda_k)\}_{p=1}^{\infty}$的收敛性, $j=0,1,\cdots,m_k-1$, $k=1,\cdots,s$,这里, $\sigma(A)=\{\lambda_1,\lambda_2,\cdots,\lambda_s\}$.

定理 1　设 $A=(a_{ij})\in\boldsymbol{M}_n(\mathbb{C})$的 s 个不同特征值 $\lambda_1,\lambda_2,\cdots,\lambda_s$ 的指标分别为 $m_1,m_2,\cdots,m_s$,且 $m=\sum\limits_{k=1}^{s}m_k$,又设 $f_1,f_2,\cdots,f_p,\cdots$为 $\mathscr{H}$内的复值函数序列. 则 $\{f_p(A)\}_{p=1}^{\infty}$收敛当且仅当 m 个数列$\{f_p^{(j)}(\lambda_k)\}_{p=1}^{\infty}$ $(0\leqslant j\leqslant m_k-1,1\leqslant k\leqslant s)$皆收敛. 此时,有 $f\in\mathscr{H}$使得

$\lim\limits_{p\to\infty}f_p^{(j)}(\lambda_k)=f^{(j)}(\lambda_k)$ $(0\leqslant j\leqslant m_k-1,1\leqslant k\leqslant s)$与$\lim\limits_{p\to\infty}f_p(A)=f(A)$.

证明　设 $A=SJS^{-1}=S\mathrm{diag}[J_1,\cdots,J_l]S^{-1}$,这里 J 为 A 的 Jordan 正规形式,诸 J_j 为 Jordan 块. 则 $f_p(A)=S\mathrm{diag}[f_p(J_1),\cdots,f_p(J_l)]S^{-1}$,其中 $f_p(J_j)$有如 3.2.1 小节(5)式的表达式, $p=1,2,\cdots$. 因此,按(1)式, $f_1(A),f_2(A),\cdots,f_p(A),\cdots$收敛当且仅当 $f_1^{(j)}(\lambda_k),f_2^{(j)}(\lambda_k),\cdots,f_p^{(j)}(\lambda_k),\cdots$收敛$(0\leqslant j\leqslant m_k-1,1\leqslant k\leqslant s)$. 若 $\lim\limits_{p\to\infty}f_p(A)$存在,则$\lim\limits_{p\to\infty}f_p^{(j)}(\lambda_k)$存在,记为 β_{jk} $(0\leqslant j\leqslant m_k-1,1\leqslant k\leqslant s)$. 选择 $f(\lambda)$为 Hermite 插值多项式满足条件: $f^{(j)}(\lambda_k)=\beta_{jk}$ $(0\leqslant j\leqslant m_k-1,1\leqslant k\leqslant s)$. 这时由 3.2.2 小节定理 1,

$$\begin{aligned}\lim_{p\to\infty}f_p(A)&=\lim_{p\to\infty}\sum_{k=1}^{s}\sum_{j=0}^{m_k-1}f_p^{(j)}(\lambda_k)E_{kj}\\&=\sum_{k=1}^{s}\sum_{j=0}^{m_k-1}f^{(j)}(\lambda_k)E_{kj}=f(A).\end{aligned}$$

□

作为定理 1 的直接推论,对于矩阵函数的级数,考虑其部分和的序列,我们也有类似的结论.

定理 2 设 $A=(a_{ij})\in \boldsymbol{M}_n(\mathbb{C})$ 同定理 1,$u_1,u_2,\cdots$ 皆为 $\mathscr{H}$ 内的复值函数. 则 $\sum\limits_{p=1}^{\infty}u_p(A)$ 收敛当且仅当 $\sum\limits_{p=1}^{\infty}u_p^{(j)}(\lambda_k)$ 对 $0\leqslant j\leqslant m_k-1,1\leqslant k\leqslant s$ 都收敛. 此时,有 $f\in\mathscr{H}$ 使得

$$\sum_{p=1}^{\infty}u_p^{(j)}(\lambda_k)=f^{(j)}(\lambda_k)(0\leqslant j\leqslant m_k-1,1\leqslant k\leqslant s)\text{ 与 }\sum_{p=1}^{\infty}u_p(A)=f(A).$$

应用定理 2 可以证明如下重要的结论,有时它可以作为某些矩阵函数的一种定义.

定理 3 设 $A\in \boldsymbol{M}_n(\mathbb{C})$ 同定理 1,且复值函数 $f(\lambda)$ 在 $\lambda=\lambda_0$ 处有 Taylor 级数展开:

$$f(\lambda)=\sum_{p=0}^{\infty}\alpha_p(\lambda-\lambda_0)^p, \tag{3}$$

此级数的收敛半径为 r. 那么,$f(A)$ 有定义且

$$f(A)=\sum_{p=0}^{\infty}\alpha_p(A-\lambda_0 I)^p \tag{4}$$

的充分与必要条件为 A 的每个不同特征值 $\lambda_1,\lambda_2,\cdots,\lambda_s$ 均满足下列两条件之一:

(1) $|\lambda_k-\lambda_0|<r$.

(2) $|\lambda_k-\lambda_0|=r$ 且对应于 $f^{(m_k-1)}(\lambda)$ 的 Taylor 级数在 $\lambda=\lambda_k$ 处收敛.

证明 设 A 的每个特征值 $\lambda_k(1\leqslant k\leqslant s)$ 满足条件(1)或(2). 则由(3)式给定的 $f(\lambda)$ 在每个 λ_k 处,$f(\lambda_k),f'(\lambda_k),\cdots,f^{(m_k-1)}(\lambda_k)$ 存在,$k=1,2,\cdots,s$,因而 $f\in\mathscr{H}$. 现令

$$u_p(\lambda)=\alpha_p(\lambda-\lambda_0)^p,\quad p=0,1,\cdots.$$

显然,$u_p\in\mathscr{H},p=0,1,\cdots$ 且在条件(1)或(2)下,数项级数 $\sum\limits_{p=0}^{\infty}u_p^{(j)}(\lambda_k)(0\leqslant j\leqslant m_k-1,1\leqslant k\leqslant s)$ 均收敛. 这时,定理 2 蕴涵 $\sum\limits_{p=0}^{\infty}u_p(A)=f(A)$,此即(4) 式. 反之,设 $f(A)$ 有定义且(4) 式成立. 则 $f^{(j)}(\lambda_k)(0\leqslant j\leqslant m_k-1,1\leqslant k\leqslant s)$ 存在,且根据定理 2,数项级数 $\sum\limits_{p=0}^{\infty}u_p^{(j)}(\lambda_k)(0\leqslant j\leqslant m_k-1,1\leqslant k\leqslant s)$ 均收敛,其中 $u_p(\lambda)=\alpha_p(\lambda-\lambda_0)^p,p=0,1,\cdots$. 因此,$A$ 的任一特征值 λ_k 必须满足 $|\lambda_k-\lambda_0|\leqslant r$. 这时,对固定 $\lambda_k\in\sigma(A)$ 来说,或有 $|\lambda_k-\lambda_0|<r$,或有 $|\lambda_k-\lambda_0|=r$ 但对应于 $f^{(m_k-1)}(\lambda)$ 的级数 $\sum\limits_{p=0}^{\infty}u_p^{(m_k-1)}(\lambda)$ 在 $\lambda=\lambda_k$ 处收敛. □

在 3.2.1 小节中,对 $A\in \boldsymbol{M}_n(\mathbb{C})$ 与 $f\in\mathscr{H}$ 定义了 $f(A)$. 上述定理表明,若 $f(\lambda)$

在 $\lambda=\lambda_0$ 处可以展开为幂级数(3),且 A 的特征值均满足条件(1)或(2),则能用矩阵级数(4)来定义 $f(A)$. 当然由于这种定义对 f 的要求更高(即 f 在 $\lambda=\lambda_0$ 邻近解析),故它不能适用于只有有限阶导数值 $f^{(j)}(\lambda_k)(0\leqslant j\leqslant m_k-1,1\leqslant k\leqslant s)$ 的函数 $f(\lambda)$ 的情形. 有幸的是,在实际应用中遇到的复值函数 f 往往满足定理 3 中的条件,因而前述的 $f(A)$ 第二种定义还是有相当的普遍性的. 特别地,如果 $f(\lambda)$ 关于复平面原点的 Taylor 级数展开 $f(\lambda)=\sum\limits_{p=0}^{\infty}\alpha_p\lambda^p$ 的收敛半径 $r=+\infty$,即 $f(\lambda)$ 为整函数,那么根据定理 3,$f(A)$ 的幂级数展开

$$f(A)=\sum_{p=0}^{\infty}\alpha_p A^p \tag{5}$$

将对所有 $A\in \boldsymbol{M}_n(\mathbb{C})$ 都收敛. 除了多项式函数以外,常见的整函数还有正弦与余弦函数、指数函数、正弦双曲与余弦双曲函数等. 根据这些复值函数在原点处的 Taylor级数展开与定理 3,对任意 $A\in \boldsymbol{M}_n(\mathbb{C})$,我们有

$$\mathrm{e}^A=\sum_{p=0}^{\infty}\frac{1}{p!}A^p,$$

$$\sin A=\sum_{p=0}^{\infty}\frac{(-1)^p}{(2p+1)!}A^{2p+1},\qquad \cos A=\sum_{p=0}^{\infty}\frac{(-1)^p}{(2p)!}A^{2p}, \tag{6}$$

$$\sinh A=\sum_{p=0}^{\infty}\frac{1}{(2p+1)!}A^{2p+1},\qquad \cosh A=\sum_{p=0}^{\infty}\frac{1}{(2p)!}A^{2p}.$$

此外,$(1-\lambda)^{-1}$ 与 $\ln(1+\lambda)$($\mathrm{Ln}(1+\lambda)$ 的主值分枝,满足 $\mathrm{Ln}1=0$)虽不是整函数,但它们在 $\lambda=0$ 处幂级数展开均有收敛半径 $r=1$,因此,若 $A\in \boldsymbol{M}_n(\mathbb{C})$ 的谱 $\sigma(A)\subset\Delta=\{\lambda\in C:|\lambda|<1\}$,则从定理 3 可推出,

$$(I-A)^{-1}=\sum_{p=0}^{\infty}A^p, \tag{7}$$

$$\ln(I+A)=\sum_{p=1}^{\infty}\frac{(-1)^{p-1}}{p}A^p.$$

(7)式中第一个式子已在第二章 2.2.2 小节定理 12 中得到.

例 4 (1) $\mathrm{e}^{A(s+t)}=\mathrm{e}^{As}\mathrm{e}^{At}$, $\forall A\in \boldsymbol{M}_n(\mathbb{C}), s,t\in\mathbb{C}$.

(2)若 $A,B\in \boldsymbol{M}_n(\mathbb{C})$ 满足 $AB=BA$,则有 $\mathrm{e}^{A+B}=\mathrm{e}^A\mathrm{e}^B$. 并且,若 $\mathrm{e}^{(A+B)t}=\mathrm{e}^{At}\mathrm{e}^{Bt}$, $\forall t\in\mathbb{C}$,则 $AB=BA$.

显然,(1)为(2)中第一个结论的特殊情形. 现设 $AB=BA$. 按(6)式,

$$\begin{aligned}\mathrm{e}^{A+B}-\mathrm{e}^A\mathrm{e}^B&=\lim_{\nu\to\infty}\left[\sum_{p=0}^{\nu}\frac{1}{p!}(A+B)^p-\sum_{q=0}^{\nu}\frac{1}{q!}A^q\sum_{k=0}^{\nu}\frac{1}{k!}B^k\right]\\&=\lim_{\nu\to\infty}\left[\sum_{p=0}^{\nu}\frac{1}{p!}\sum_{j=0}^{p}\binom{p}{j}A^jB^{p-j}-\sum_{q=0}^{\nu}\sum_{k=0}^{\nu}\frac{1}{q!k!}A^qB^k\right],\end{aligned}$$

式中，$\binom{p}{j}=\dfrac{p!}{j!(p-j)!}$. 我们在上式推导中用到事实：若矩阵级数 $\sum_p A_p$ 与 $\sum_s B_s$ 绝对收敛，则 $(\sum_p A_p)(\sum_s B_s)=\sum_{p,s} A_pB_s$（见第二章 2.2.2 小节）. 在上式的最后方括号中，对第一个和项做变换：$q=j$ 与 $k=p-j$，它变为 $\sum\limits_{\substack{q,k\geqslant 0\\ q+k\leqslant \nu}} \dfrac{1}{q!k!}A^qB^k$. 于是对于 $q+k\leqslant \nu$，A^qB^k 同时出现在前后两个和项里，因而这些项相抵消去，只余下

$$-\sum_q\sum_k \frac{1}{q!k!}A^qB^k, \tag{8}$$

其中，$q\leqslant\nu, k\leqslant\nu$ 且 $q+k>\nu$. 由于 $q+k>\nu$ 蕴涵 q 与 k 两个整数之中至少有一大于 $\nu/2$ 的整数部分，记为 $[\nu/2]$，故对 $\boldsymbol{M}_n(\mathbb{C})$ 上任一导出的矩阵范数 $\|\cdot\|$ 有

$$\begin{aligned}\left\|\sum_q\sum_k \frac{1}{q!k!}A^qB^k\right\| &\leqslant \sum_{q=0}^{\infty}\sum_{k=[\nu/2]}^{\infty}\frac{1}{q!k!}\|A\|^q\|B\|^k\\ &\quad+\sum_{k=0}^{\infty}\sum_{q=[\nu/2]}^{\infty}\frac{1}{q!k!}\|A\|^q\|B\|^k\\ &=\left(\sum_{q=0}^{\infty}\frac{1}{q!}\|A\|^q\right)\left(\sum_{k=[\nu/2]}^{\infty}\frac{1}{k!}\|B\|^k\right)+\left(\sum_{k=0}^{\infty}\frac{1}{k!}\|B\|^k\right)\left(\sum_{q=[\nu/2]}^{\infty}\frac{1}{q!}\|A\|^q\right)\\ &=e^{\|A\|}\sum_{k=[\nu/2]}^{\infty}\frac{1}{k!}\|B\|^k+e^{\|B\|}\sum_{q=[\nu/2]}^{\infty}\frac{1}{q!}\|A\|^q.\end{aligned}$$

但是，$\lim\limits_{\nu\to\infty}\sum\limits_{k=[\nu/2]}^{\infty}\dfrac{1}{k!}\|B\|^k=\lim\limits_{\nu\to\infty}\sum\limits_{q=[\nu/2]}^{\infty}\dfrac{1}{q!}\|A\|^q=0$，因而由上述不等式推出(8)式中和项当 $\nu\to\infty$ 时趋于 n 阶零矩阵. 因此，$e^{A+B}=e^Ae^B$.

现在证明(2)中后一结论. 按(6)式我们有

$$e^{(A+B)t}=\sum_{p=0}^{\infty}\frac{1}{p!}(A+B)^pt^p,$$

$$e^{At}e^{Bt}=\sum_{q=0}^{\infty}\frac{1}{q!}A^qt^q\sum_{k=0}^{\infty}\frac{1}{k!}B^kt^k.$$

$e^{(A+B)t}$ 展开中前三项为 $I+(A+B)t+\dfrac{1}{2}(A+B)^2t^2$. 对于 $e^{At}e^{Bt}$，我们有

$$\begin{aligned}e^{At}e^{Bt}&=I+(A+B)t+\frac{1}{2}(A^2+2AB+B^2)t^2+\frac{1}{2}(AB^2+A^2B)t^3\\ &\quad+\left(I+At+\frac{1}{2}A^2t^2\right)\sum_{k=3}^{\infty}\frac{1}{k!}B^kt^k+\left(\sum_{q=3}^{\infty}\frac{1}{q!}A^qt^q\right)\left(I+Bt+\frac{1}{2}B^2t^2\right)\\ &\quad+\left(\sum_{q=3}^{\infty}\frac{1}{q!}A^qt^q\right)\left(\sum_{k=3}^{\infty}\frac{1}{k!}B^kt^k\right).\end{aligned}\tag{9}$$

因此，$e^{(A+B)t}=e^{At}e^{Bt}$，$\forall t\in\mathbb{C}$ 蕴涵

$$\begin{aligned}&\frac{1}{2}(A+B)^2t^2+\sum_{q=3}^{\infty}\frac{1}{q!}(A+B)^qt^q\\&=\frac{1}{2}(A^2+2AB+B^2)t^2+\frac{1}{2}(AB^2+A^2B)t^3+\cdots,\end{aligned}$$

式中,“…”表示(9)式右端最后三项之和. 对任取的 $t\neq 0$,有

$$\frac{1}{2}(A+B)^2+\sum_{q=3}^{\infty}\frac{1}{q!}(A+B)^qt^{q-2}=\frac{1}{2}(A^2+2AB+B^2)+\frac{1}{2}(AB^2+A^2B)t+\cdots. \tag{10}$$

让 $t\to 0$,由于

$$\begin{aligned}\left\|\sum_{q=3}^{\infty}\frac{1}{q!}(A+B)^qt^{q-2}\right\|&\leqslant\sum_{q=3}^{\infty}\frac{1}{q!}\|A+B\|^q\,|t|^{q-2}\\&\leqslant|t|\sum_{q=0}^{\infty}\frac{1}{q!}\|A+B\|^q=|t|\,e^{\|A+B\|},\end{aligned}$$

故(10)式左端第二项趋于 O. 同理可证当 $t\to 0$ 时(10)式右端除了第一项外,其他各项都趋于 O. 因此,

$$\frac{1}{2}(A+B)^2=\frac{1}{2}(A^2+2AB+B^2),$$

此直接推出 $AB+BA=2AB$,因而 $AB=BA$. □

例 5 设

$$A=\begin{bmatrix}0&1\\1&0\end{bmatrix},\quad B=\begin{bmatrix}0&1\\0&\varepsilon\end{bmatrix},\quad C=\begin{bmatrix}0&1\\-1&0\end{bmatrix}.$$

求 e^{At}, e^{Bt} 与 e^{Ct}, $\forall t\in\mathbb{C}$.

容易验证: $A^{2p+1}=A$, $A^{2p}=I_2$, $p=0,1,\cdots$. 按(6)式, $e^{At}=\sum_{p=0}^{\infty}\frac{1}{p!}A^pt^p=\sum_{p=0}^{\infty}\frac{t^{2p+1}}{(2p+1)!}A+\sum_{p=0}^{\infty}\frac{t^{2p}}{(2p)!}I_2=(\sinh t)A+(\cosh t)I_2$. 因此,

$$e^{At}=\begin{bmatrix}\cosh t&\sinh t\\\sinh t&\cosh t\end{bmatrix}.$$

接着,由于 $B^p=\varepsilon^{p-1}B$, $p=1,2,\cdots$,故我们有 $e^{Bt}=\sum_{p=0}^{\infty}\frac{1}{p!}B^pt^p=I_2+\sum_{p=1}^{\infty}\frac{\varepsilon^{p-1}t^p}{p!}B=I_2+\varepsilon^{-1}(e^{\varepsilon t}-1)B$. 因此,

$$e^{Bt}=\begin{bmatrix}1&\varepsilon^{-1}(e^{\varepsilon t}-1)\\0&e^{\varepsilon t}\end{bmatrix}.$$

最后,由于 $C^{2p}=(-1)^pI_2$, $C^{2p+1}=(-1)^pC$, $p=0,1,\cdots$,故有

$$e^{\mathcal{C}} = \sum_{p=0}^{\infty} \frac{1}{p!} C^p t^p = \sum_{p=0}^{\infty} \frac{(-1)^p t^{2p+1}}{(2p+1)!} C + \sum_{p=0}^{\infty} \frac{(-1)^p t^{2p}}{(2p)!} I_2 = (\sin t) C + (\cos t) I_2.$$

因此,

$$e^{\alpha} = \begin{bmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{bmatrix}.$$

我们顺便指出,若取 $D = -\mathrm{i}Ct, \mathrm{i}^2 = -1$,则 $D = \begin{bmatrix} 0 & -\mathrm{i}t \\ \mathrm{i}t & 0 \end{bmatrix}$. 当 $t \in \mathbb{C}$ 的虚部$\neq 0$ 时 D 不是 Hermite 矩阵,但 $e^{\mathrm{i}D} = e^{\mathcal{C}}$ 为酉矩阵.(注意联系 3.1.2 小节命题 4) □

例 6 设 $A = (a_{ij}) \in \boldsymbol{M}_n(\mathbb{C})$ 为 Hermite 正定矩阵. 讨论矩阵 $K = (k_{ij}) \equiv (e^{a_{ij}}) \in \boldsymbol{M}_n(\mathbb{C})$ 的正定性.

由于

$$e^{a_{ij}} = 1 + a_{ij} + \frac{1}{2!} a_{ij}^2 + \frac{1}{3!} a_{ij}^3 + \cdots, \qquad \forall i, j,$$

我们有

$$K = (e^{a_{ij}}) = J + A + \frac{A \circ A}{2!} + \frac{A \circ A \circ A}{3!} + \cdots, \tag{11}$$

式中,J 表示元素均为 1 的 n 阶矩阵,$A \circ B$ 表示矩阵 A 与 B 的 Hadamard 积(见第一章 1.2.2 小节). 按第一章 1.2.2 小节定理 13 及其后面的说明,(11)式中矩阵级数除了 $J \geqslant O$ 外,其他项都是 Hermite 正定矩阵. 记(11)式中矩阵级数的部分和为 S_m,则对 $m \geqslant 2$, S_m 为 Hermite 正定的,因而对任意 $\boldsymbol{0} \neq \boldsymbol{x} \in \mathbb{C}^n$,内积 $\langle K\boldsymbol{x}, \boldsymbol{x} \rangle = \lim_{m \to \infty} \langle S_m \boldsymbol{x}, \boldsymbol{x} \rangle \geqslant \langle A\boldsymbol{x}, \boldsymbol{x} \rangle > 0$,即有 $K > O$. □

顺便指出,前例中矩阵 K 不是 A 的指数函数 e^A.(虽然,AHermite 正定也蕴涵 e^A 是 Hermite 正定的)通常,称这里的 K 为 A 的 **Schur 指数函数**.

例 7 设 $A \in \boldsymbol{M}_n(\mathbb{C})$,且 $B_m(t) \in \boldsymbol{M}_n(\mathbb{C})(m = 1, 2 \cdots)$ 满足 $\lim_{m \to \infty} \| B_m(t) \| = 0$ 对 $t \in \mathscr{D}$ 一致成立,这里 $\mathscr{D} \subset \mathbb{C}$ 为有界区域. 则对 $t \in \mathscr{D}$ 下式一致成立:

$$\lim_{m \to \infty} [I + (t/m)(A + B_m(t))]^m = e^{At}. \tag{12}$$

特殊地,我们有 $\lim_{m \to \infty} [I + (t/m)A]^m = \lim_{m \to \infty} [I - (t/m)A]^{-m} = e^{At}$.(显然,这是数值情形公式 $\lim_{m \to \infty} [1 + \alpha/m]^m = e^{\alpha}$ 的推广)

令 $V_m(t) = I + (t/m)(A + B_m(t)), m = 1, 2, \cdots$. 对 $\boldsymbol{M}_n(\mathbb{C})$ 上任意导出的矩阵范数 $\| \cdot \|$ 与 $0 \leqslant k \leqslant m$,按题设有

$$\begin{aligned} \| V_m^k(t) \| &\leqslant [1 + | t/m | (\| A \| + \| B_m(t) \|)]^k \\ &\leqslant [1 + | t/m | (\| A \| + \| B_m(t) \|)]^m \\ &\leqslant e^{(\| A \| + \| B_m(t) \|)|t|} \leqslant M, \qquad \forall t \in \mathscr{D} \end{aligned}$$

其中,M 为与 $t \in \mathscr{D}$ 无关的某个常数. 现令 $U_m(t) = e^{(t/m)A}$. 则 $U_m^m(t) = e^{At}$,且也有

$\| U_m^k(t) \| \leqslant e^{\|A\| \ |t|} \leqslant M$, $\forall t \in \mathscr{D}, k=0,1,\cdots,m$. 按第一章习题 1.1.1 题 4，

$$V_m^m(t) - U_m^m(t) = \sum_{j=0}^{m-1} V_m^j(t)[V_m(t) - U_m(t)]U_m^{m-1-j}(t).$$

但是，

$$\begin{aligned} V_m(t) - U_m(t) &= I + (t/m)[A + B_m(t)] - [I + (t/m)A + t^2/(2m^2)A^2 + \cdots] \\ &= t/m[B_m(t) - t/(2m)A^2 + \cdots], \end{aligned}$$

因而对充分大的 m，我们有

$$\| V_m(t) - U_m(t) \| \leqslant | t/m | \ \| B_m(t) \| + | t/m |^2 \| A \|^2, \qquad \forall t \in \mathscr{D}.$$

这时，(12)式由以下估计式推出：对充分大的 m 与 $t \in \mathscr{D}$，

$$\| V_m^m(t) - U_m^m(t) \| \leqslant mM^2[| t/m | \ \| B_m(t) \| + | t/m |^2 \| A \|^2],$$

因而 $\lim\limits_{m\to\infty} \| V_m^m(t) - U_m^m(t) \| = 0$ 对 $t \in \mathscr{D}$ 一致成立.

特殊地，如分别地取 $B_m(t) \equiv O$ 与 $B_m(t) \equiv \frac{t}{m}A^2 + \frac{t^2}{m^2}A^3 + \cdots, t \in \mathscr{D}$，并应用事实：$[I - (t/m)A]^{-1} = \sum\limits_{j=0}^{\infty}(t/m)^j A^j$，则(12)式变为

$$e^{At} = \lim_{m\to\infty}[I + (t/m)A]^m = \lim_{m\to\infty}[I - (t/m)A]^{-m},$$

它在复参数 t 的任一有界区域内一致成立. □

习题 3.2.3

1. 试应用 3.2.2 小节定理 1 证明本小节的定理 1.

2. 设 $f(\lambda) = \arctan\lambda = \sum\limits_{p=0}^{\infty} \frac{(-1)^p}{2p+1}\lambda^{2p+1}$. 试证：若 $A \in \boldsymbol{M}_n(\mathbb{C})$ 有 $\rho(A) < 1$，则 $\arctan A = \sum\limits_{p=0}^{\infty} \frac{(-1)^p}{2p+1}A^{2p+1}$.

3. 设 $A, B \in \boldsymbol{M}_n(\mathbb{C})$ 为 Hermite 矩阵. 试证：

$$\mathrm{tr}(e^{A+B}) \leqslant \mathrm{tr}(e^A e^B).$$

4. 设

$$A = \begin{bmatrix} 3 & 1 & 1 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{bmatrix}.$$

试求 e^A.

5. (1) 设 $\hat{J}$ 为对应于特征值 λ 的 Jordan 块，阶数为 p，又设 $\mu = \mathrm{Re}(\lambda)$ 与 $f(t) = t^{p-1}e^{\mu t}$. 试证：存在正常数 k_1 与 k_2 使得

$$k_1 f(t) \leqslant \| e^{\hat{J}t} \|_2 \leqslant k_2 f(t), \qquad \forall t \geqslant 1.$$

(2) 设 $A \in \boldsymbol{M}_n(\mathbb{C})$，$\mu = \max\limits_{\lambda\in\sigma(A)} \mathrm{Re}(\lambda)$，且 p 为满足 $\mu = \mathrm{Re}(\lambda)$ 的 A 特征值 λ 的最大指标. 试证：存在正常数 R, K_1 与 K_2 使得

$$K_1 f(t) \leqslant \| e^{At} \|_2 \leqslant K_2 f(t), \qquad \forall t \geqslant R.$$

(3)应用(2)的结论证明：

(i)当且仅当 $\mu<0$ 时，$\lim\limits_{t\to\infty}\|e^{At}\|_2=0$.

(ii)若 $\mu\leqslant 0$ 与 $p=1$，则 $\|e^{At}\|_2\leqslant K_2,\forall t\geqslant R$.

(iii)若 $\mu<0$ 则有正常数 K 与 α 使得 $\|e^{At}\|_2\leqslant Ke^{-\alpha t},\forall t\geqslant R$.

6. 设 $A,B\in \boldsymbol{M}_n(\mathbb{C})$. 试证：

$$\lim_{m\to\infty}[e^{(t/m)A}e^{(t/m)B}]^m = e^{(A+B)t}$$

在复参数 t 的任一有界区域内一致成立(**Lie-Trotter 乘积公式**). 并且，如将上式左端中 $e^{(t/m)A}$ 换成 $I+(t/m)A$ 或 $(I-(t/m)A)^{-1}$ 或 $[I-(t/m)(A/k)]^{-k}$，或更一般地，换成 $\varphi(tA/m)$，这里 $\varphi(\xi)$ 为满足 $\varphi(0)=\varphi'(0)=1$ 且在 $\xi=0$ 处解析的任意复值函数，则原式仍成立.

3.3 矩阵函数 $f(A)$：f 为解析函数情形

在前面两节中，讨论了 $\boldsymbol{M}_n(\mathbb{C})$ 中简单矩阵与一般矩阵 A 的函数 $f(A)$. 若 A 的 s 个不同特征值 $\lambda_1,\cdots,\lambda_s$ 分别有指标 $m_1,\cdots,m_s$，$m=\sum\limits_{k=1}^{s}m_k$，则只要 $f\in\mathscr{H}$，即 $f^{(j)}(\lambda_k)$ 有定义 $(0\leqslant j\leqslant m_k-1,1\leqslant k\leqslant s)$，便可定义 $f(A)$. 显见，这种 $f(A)$ 的定义虽对 f 要求较低，但直接有赖于 A 本身的代数结构，即 Jordan 正规形式的结构，因而不便于推广到无限维空间的情形. 其次，这种定义与其说给出矩阵函数不如说给出 $f(A)$ 的具体涵义，因为我们很少涉及自变量 A 变化的情况. 值得注意的是，在 3.2.3 小节定理 3 及其后面说明中，我们对在复平面某点 $\lambda=\lambda_0$ 邻近解析的复值函数 f，定义了 $f(A)$. 这种定义比较容易推广到无限维空间的情形. 但由于它与 $f(\lambda)$ 的 Taylor 级数展开联系在一起，讨论起来多有不便之处. 众所周知，在复分析中，一般考虑用 Cauchy 积分表示解析函数. 自然地，我们也希望用对应的矩阵 Cauchy 定理来定义 $f(A)$，并讨论它的有关性质. 这些内容可以自然地推广到无限维空间的情形，因而有很高的理论价值(见文献[2]).

在正式讨论之前，先引入一些预备知识.

3.3.1 矩阵值函数的分析运算与矩阵的预解式

带有实或复变量 t 并取值在 $\boldsymbol{M}_{n,m}(\mathbb{C})$ 内的映射 $A_t=A(t)$ 称为**矩阵值函数**. 特殊地当 $m=1$ 时，我们得到**向量值函数**，通常用 $\boldsymbol{u}_t=\boldsymbol{u}(t)$ 等形式表示. 今后主要研究具有方阵值(即 $m=n$ 的情形)的矩阵值函数与向量值函数(即 $m=1$ 的情形).

先考虑向量值函数 $\boldsymbol{u}_t=\boldsymbol{u}(t)$，$t\in\mathscr{D}$，其中 $\mathscr{D}$ 为 $\mathbb{R}$ 或 $\mathbb{C}$ 内开集(如开集 $\mathscr{D}\subset\mathbb{R}$，则 $\mathscr{D}$ 为开区间的并集，如 $\mathscr{D}$ 还是单连通的，则它为开区间). 极限关系 $\lim\limits_{t\to t_0}\boldsymbol{u}(t)=\boldsymbol{v}$ 由 $\lim\limits_{t\to t_0}\|\boldsymbol{u}(t)-\boldsymbol{v}\|=0$ 来定义，式中 $\|\cdot\|$ 为 $\boldsymbol{u}_t$ 值域空间 $\mathbb{C}^n$ 上任一向量范数，$t\neq t_0$.

$\boldsymbol{u}_t$ 在 $t=t_0\in\mathscr{D}$ 处**连续**，假如 $\lim\limits_{t\to t_0}\boldsymbol{u}(t)=\boldsymbol{u}(t_0)$. 若 $\boldsymbol{u}_t$ 在 $\mathscr{D}$ 每点处连续，则称 $\boldsymbol{u}_t$ **在 $\mathscr{D}$ 内连续**. 当 $\boldsymbol{u}(t)=(u_1(t),\cdots,u_n(t))^{\mathrm{T}}$ 时，上述连续性定义显然等价于所有 $u_j(t)$ 在 $t=t_0$ 或在 $\mathscr{D}$ 内的连续性. $\boldsymbol{u}(t)$ 的**导数**由

$$\boldsymbol{u}'(t)=\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{u}(t)=\lim_{h\to 0}h^{-1}(\boldsymbol{u}(t+h)-\boldsymbol{u}(t)) \tag{1}$$

给出，只要这个极限存在. 显然，$\boldsymbol{u}'(t)=(u'_1(t),\cdots,u'_n(t))^{\mathrm{T}}$. 如同数值函数情形，我们还可以引入积分的定义并讨论导数与积分的公式，例如，有

$$\frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{u}(t)+\boldsymbol{v}(t))=\boldsymbol{u}'(t)+\boldsymbol{v}'(t),$$

$$\frac{\mathrm{d}}{\mathrm{d}t}(\phi(t)\boldsymbol{u}(t))=\phi(t)\boldsymbol{u}'(t)+\phi'(t)\boldsymbol{u}(t), \tag{2}$$

式中，$\phi(t)$ 为数值函数. 虽然当自变量 t 为实或复数时，向量值函数 $\boldsymbol{u}(t)$ 的导数形式上没有什么区别，但类似于数值函数情形，这二者之间有着本质的不同. 当 $\boldsymbol{u}(t)$ 在复平面 $\mathbb{C}$ 的区域 $\mathscr{D}$ 内有定义且处处有一阶导数时，则称 $\boldsymbol{u}(t)$ 是在 $\mathscr{D}$ 内**解析**的(analytic)或**全纯**的(holomorphic). 绝大部分的复解析函数结果可用于全纯的向量值函数，特别有 Cauchy 积分定理，Taylor 与 Laurent 展开以及 Liouville 定理，等等. 例如，若 $t=0$ 为全纯函数 $\boldsymbol{u}(t)$ 的**孤立奇点**，则

$$\boldsymbol{u}(t)=\sum_{j\to-\infty}^{+\infty}a_jt^j,\quad a_j=\frac{1}{2\pi\mathrm{i}}\int_{\mathscr{L}}t^{-j-1}\boldsymbol{u}(t)\mathrm{d}t, \tag{3}$$

式中，$\mathscr{L}$ 为围绕 $t=0$ 的正向封闭曲线. 当 $a_j=0,\forall j<0$ 时，则 $t=0$ 为**可去奇点**；当 $a_{-k}\neq 0$ 而 $a_j=0,\forall j<-k(<0)$ 时，则 $t=0$ 为 **k 阶极点**；余下情形 $t=0$ 为**本性奇点**.

对于矩阵值函数情形，也有类似于刚才对于向量值函数的结果，其中向量范数应换成矩阵范数，它满足 $\|I\|=1$. 但有一些特殊的地方，如当 $A(t)$ 与 $B(t)$ 为矩阵值函数时，可以讨论 $A(t)B(t)$ 与 $A(t)\boldsymbol{u}(t)$ 的连续性、可微性与可积性. 例如，只要下面诸项有意义且右端诸导数存在，则有

$$\frac{\mathrm{d}}{\mathrm{d}t}(A(t)\boldsymbol{u}(t))=A'(t)\boldsymbol{u}(t)+A(t)\boldsymbol{u}'(t), \tag{4}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}(A(t)B(t))=A'(t)B(t)+A(t)B'(t).$$

此外，当 $A(t)^{-1}$ 与 $A'(t)$ 存在时，便有

$$\frac{\mathrm{d}}{\mathrm{d}t}A(t)^{-1}=-A(t)^{-1}\frac{\mathrm{d}}{\mathrm{d}t}A(t)A(t)^{-1}. \tag{5}$$

事实上，当 $A(t)^{-1}$ 存在，且 $h\in\mathscr{D}$ 满足 $\|A(t+h)-A(t)\|<\|A(t)^{-1}\|^{-1}/2$ 时，$A(t+h)=A(t+h)-A(t)+A(t)=A(t)(I+A(t)^{-1}(A(t+h)-A(t)))$，其中由于 $\|A(t)^{-1}(A(t+h)-A(t))\|\leqslant\|A(t)^{-1}\|\ \|A(t+h)-A(t)\|<1/2$，故按第

二章 2.2.2 小节定理 12，$I+A(t)^{-1}(A(t+h)-A(t))$为 $\boldsymbol{M}_n(\mathbb{C})$中非奇异矩阵，于是，$A(t+h)$为两个可逆矩阵之积也是可逆的. 此时，

$$
\begin{aligned}
&\| A(t+h)^{-1} - A(t)^{-1} + A(t)^{-1}(A(t+h) - A(t))A(t)^{-1} \| \\
&= \| [A(t+h)^{-1}A(t) - I + A(t)^{-1}(A(t+h) - A(t))]A(t)^{-1} \| \\
&= \| \sum_{k=2}^{\infty} (-1)^k [A(t)^{-1}(A(t+h) - A(t))]^k A(t)^{-1} \| \\
&\leqslant \frac{\| A(t)^{-1}(A(t+h) - A(t)) \|^2}{1 - \| A(t)^{-1}(A(t+h) - A(t)) \|} \| A(t)^{-1} \| \\
&\leqslant 2 \| A(t)^{-1} \|^3 \| A(t+h) - A(t) \|^2.
\end{aligned}
\tag{6}
$$

这表明，$\lim\limits_{h\to 0}A(t+h)^{-1}=A(t)^{-1}$，即 $A(t)^{-1}$在 t 处连续，并且，用$|1/h|$乘以(6)式各项便知，(5)式成立. 这样一来，若 $A(t)$解析，且 $A(t)^{-1}$存在，则 $A(t)^{-1}$也解析.

应用上述证明可顺便地得到，$\boldsymbol{M}_n(\mathbb{C})$中所有非奇异矩阵的集合为 $\boldsymbol{M}_n(\mathbb{C})$的开集.

我们还可以定义矩阵值函数 $A(t)$的高阶导数$\frac{\mathrm{d}^r}{\mathrm{d}t^r}A(t)$或 $A^{(r)}(t)$，$r=1,2,\cdots$，并且，数值函数高阶导数的许多法则在矩阵值函数中也有它们对应的结果. 例如，设 $R_\lambda\equiv(\lambda I-A)^{-1}$，$\lambda\notin\sigma(A)$，$A\in\boldsymbol{M}_n(\mathbb{C})$给定. 则按(4)与(5)式有

$$
\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}\lambda}R_\lambda = -R_\lambda^2, \qquad \frac{\mathrm{d}^2}{\mathrm{d}\lambda^2}R_\lambda = 2R_\lambda^3, \cdots, \\
&\frac{\mathrm{d}^r}{\mathrm{d}\lambda^r}R_\lambda = (-1)^r r! R_\lambda^{r+1}, \qquad r = 1,2,\cdots.
\end{aligned}
\tag{7}
$$

矩阵值函数 $A(t)$的积分也可仿照数值函数的情形来定义. 例如，假定 $A(t)$为区间$[a,b]$上关于实变量 t 的连续函数，Riemann 积分 $\int_a^b A(t)\mathrm{d}t$ 可以定义为和 $\sum (t_j - t_{j-1})A(t_j)$ 的合适极限，其中，$a = t_0 < t_1 < \cdots < t_m = b$ 为积分区间$[a,b]$ 的一种分划. 类似地，对复变量 t 的连续函数 $A(t)$ 与有向曲线 Γ 可以定义 $\int_\Gamma A(t)\mathrm{d}t$. 显然，若 $A(t) = (a_{ij}(t))$，则我们有$\int_\Gamma A(t)\mathrm{d}t = \left(\int_\Gamma a_{ij}(t)\mathrm{d}t\right)_{i,j=1}^n$，即它为各元素 $a_{ij}(t)$ 积分的矩阵. 一些常用的积分规则仍然有效，如

$$
\begin{aligned}
&\int(\alpha A(t) + \beta B(t))\mathrm{d}t = \alpha\int A(t)\mathrm{d}t + \beta\int B(t)\mathrm{d}t, \\
&\left\|\int A(t)\mathrm{d}t\right\| \leqslant \int \| A(t) \| \mathrm{d}t, \\
&\int S\boldsymbol{u}(t)\mathrm{d}t = S\int \boldsymbol{u}(t)\mathrm{d}t, \\
&\int A(t)\boldsymbol{u}\mathrm{d}t = \left(\int A(t)\mathrm{d}t\right)\boldsymbol{u},
\end{aligned}
\tag{8}
$$

$$\int SA(t)\mathrm{d}t = S\int A(t)\mathrm{d}t,$$

$$\int A(t)S\mathrm{d}t = \left(\int A(t)\mathrm{d}t\right)S,$$

式中，$S\in \boldsymbol{M}_n(\mathbb{C})$；$\alpha,\beta\in\mathbb{C}$；$\boldsymbol{u}(t)$为向量值函数（其值域在$\mathbb{C}^n$内），$\boldsymbol{u}\in\mathbb{C}^n$；$\|\cdot\|$为$\boldsymbol{M}_n(\mathbb{C})$上某导出的矩阵范数.

从前面看出，矩阵值函数$A(t)$的连续性、可微性与可积性如同向量值函数情形都是在范数拓扑意义下讨论的，因此分别等价于$A(t)$诸元素$a_{ij}(t)$的相应的性质. 由于这个基本原因，复解析函数论中许多结果稍加修正后则可适用于矩阵值函数，其中特别有 Cauchy 积分定理，均值公式，Taylor 与 Laurent 展开以及 Liouville 定理等.

现在讨论矩阵$A\in\boldsymbol{M}_n(\mathbb{C})$的**预解式**，即$R_\lambda\equiv(\lambda I-A)^{-1}$. 由(7)式看出，$R_\lambda$为$\mathbb{C}\backslash\sigma(A)$上的全纯矩阵值函数. 通常，集合$\mathbb{C}\backslash\sigma(A)$叫作$A$的**预解集**，记为$\mathscr{R}(A)$.

A的预解式R_λ在$\mathscr{R}(A)$内满足如下的所谓**第一预解式方程**：

$$R_{\lambda_1}-R_{\lambda_2}=(\lambda_2-\lambda_1)R_{\lambda_1}R_{\lambda_2},\ \forall\lambda_1,\lambda_2\in\mathscr{R}(A). \tag{9}$$

（这是数值公式$(\lambda_1-\alpha)^{-1}-(\lambda_2-\alpha)^{-1}=(\lambda_2-\lambda_1)(\lambda_1-\alpha)^{-1}(\lambda_2-\alpha)^{-1}$的矩阵情形推广）事实上，(9)式左端等于$R_{\lambda_1}(\lambda_2 I-A)R_{\lambda_2}-R_{\lambda_1}(\lambda_1 I-A)R_{\lambda_2}=(\lambda_2-\lambda_1)R_{\lambda_1}R_{\lambda_2}$. (9)式蕴涵$R_{\lambda_1}$与$R_{\lambda_2}$可交换，且

$$R_{\lambda_1}=(I-(\lambda_1-\lambda_2)R_{\lambda_1})R_{\lambda_2}=(I-(\lambda_2-\lambda_1)R_{\lambda_2})^{-1}R_{\lambda_2}.$$

于是，当给定$\lambda_0\in\mathscr{R}(A)$且$|\lambda-\lambda_0|<\|R_{\lambda_0}\|^{-1}$时（$\|\cdot\|$为$\boldsymbol{M}_n(\mathbb{C})$上某导出的矩阵范数），$\lambda\in\mathscr{R}(A)$，并且，

$$R_\lambda=(I-(\lambda_0-\lambda)R_{\lambda_0})^{-1}R_{\lambda_0}=\sum_{j=0}^{\infty}(\lambda_0-\lambda)^j R_{\lambda_0}^{j+1}. \tag{10}$$

(10)式中级数对　$|\lambda-\lambda_0|<(\limsup\limits_{j\to\infty}\|R_{\lambda_0}^j\|^{1/j})^{-1}=(\rho(R_{\lambda_0}))^{-1}$是绝对收敛的，即

数项级数$\sum\limits_{j=0}^{\infty}|\lambda_0-\lambda|^j\|R_{\lambda_0}\|^{j+1}$当$|\lambda-\lambda_0|<(\rho(R_{\lambda_0}))^{-1}$时收敛. 因此，$R_\lambda$在$A$的预解集内解析.

对充分大$|\lambda|$，譬如$|\lambda|>\|A\|$，$R_\lambda=(\lambda I-A)^{-1}$有定义且有展开式

$$R_\lambda=\lambda^{-1}(I-\lambda^{-1}A)^{-1}=\sum_{j=0}^{\infty}\lambda^{-j-1}A^j.$$

上式中级数当$|\lambda|>\rho(A)$时绝对收敛，因而R_λ在$\lambda=\infty$处为全纯的.

$A\in\boldsymbol{M}_n(\mathbb{C})$的预解式$R_\lambda$为$\lambda$的有理矩阵值函数，它在$A$的谱$\sigma(A)$元素$\lambda_1$，$\lambda_2,\cdots,\lambda_s$处有奇异性（都是$R_\lambda$的极点）. 现在考虑这些$R_\lambda$极点的阶.

如前假定A有最小多项式$m(\lambda)=(\lambda-\lambda_1)^{m_1}(\lambda-\lambda_2)^{m_2}\cdots(\lambda-\lambda_s)^{m_s}$. 则当$\lambda\neq\mu$时，有复数$\alpha_{ij}(0\leqslant i,j\leqslant m-1,m=m_1+\cdots+m_s)$使得

$$\frac{m(\lambda)-m(\mu)}{\lambda-\mu}=\sum_{j=0}^{m-1}\left(\sum_{i=0}^{m-1}\alpha_{ij}\lambda^{i}\right)\mu^{j} \tag{11}$$

成立. 取定$\lambda\in\mathscr{R}(A)$,用A代换上式中的μ,则得

$$(m(\lambda)I-m(A))R_{\lambda}=\sum_{j=0}^{m-1}\left(\sum_{i=0}^{m-1}\alpha_{ij}\lambda^{i}\right)A^{j}.$$

由于$m(A)=O$与$m(\lambda)\neq 0$,故上式可以变换为

$$R_{\lambda}=\frac{1}{m(\lambda)}\sum_{j=0}^{m-1}\left(\sum_{i=0}^{m-1}\alpha_{ij}\lambda^{i}\right)A^{j}. \tag{12}$$

因为$\sum_{i=0}^{m-1}\alpha_{ij}\lambda^{i}\in\mathscr{P}_{m-1}$,$m(\lambda)$为$\lambda$的$m$次多项式,所以(12)式蕴涵$R_{\infty}=O$. 同时,对任意$\lambda_k\in\sigma(A)$,$\sum_{i=0}^{m-1}\alpha_{ij}\lambda_k^{i}$至少对某个$j\,(0\leqslant j\leqslant m-1)$不为零.(否则,从(11)式看出,对任意$\mu\neq\lambda_k$都有$m(\lambda_k)-m(\mu)=0$,即有$m(\mu)\equiv 0$,推出矛盾). 因此,由(12)式看出,$\lambda_k\in\sigma(A)$为$R_{\lambda}$的$m_k$阶的极点. 这表明$R_{\lambda}$是一个亚纯函数.

这样一来,我们证明了下面定理.

定理 1 设$A\in\boldsymbol{M}_n(\mathbb{C})$有最小多项式$m(\lambda)=(\lambda-\lambda_1)^{m_1}\cdots(\lambda-\lambda_s)^{m_s}$. 则$A$的预解式$R_{\lambda}=(\lambda I-A)^{-1}$是$\lambda$的有理矩阵值函数,它在$\lambda_k\in\sigma(A)$处有阶数为$m_k$的极点,$k=1,\cdots,s$,并且$R_{\infty}=O$.

我们指出,A的预解式在矩阵函数的积分形式定义中(见 3.3.2 小节)与在矩阵特征值解析扰动问题中都起着重要的作用.

习题 3.3.1

1. 设$A\in\boldsymbol{M}_n(\mathbb{C})$. 试证:$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{At}=A\mathrm{e}^{At}=\mathrm{e}^{At}A$.

2. 设$A(t)$对t可微. 试证:$\frac{\mathrm{d}}{\mathrm{d}t}(A(t))^2=A'(t)A(t)+A(t)A'(t)$. 并且一般地,我们有$\frac{\mathrm{d}}{\mathrm{d}t}(A(t))^p=\sum_{j=1}^{p}A^{j-1}(t)A'(t)A^{p-j}(t),\ p=1,2,\cdots$.

3. 设$A(t)$对t可微,且$A(t)^{-1}$存在. 试证:

$$\frac{\mathrm{d}}{\mathrm{d}t}(A(t))^{-p}=-(A(t))^{-p}\frac{\mathrm{d}}{\mathrm{d}t}(A(t))^{p}(A(t))^{-p},\qquad p=1,2,\cdots.$$

4. 设$H(t)=(h_{ij}(t))\in\boldsymbol{M}_n(\mathbb{C})$为给定的实变量$t$的可微的 Hermite 矩阵值函数,对每个$t\in\mathbb{R}$,$H(t)$有特征值$\lambda_1(t),\lambda_2(t),\cdots,\lambda_n(t)$(计入重根),且$H(t)=U(t)\Lambda(t)U^*(t)$,其中$U(t)=(\boldsymbol{u}_1(t)\ \cdots\ \boldsymbol{u}_n(t))\in\boldsymbol{M}_n(\mathbb{C})$为酉矩阵,$\Lambda(t)=\mathrm{diag}[\lambda_1(t),\cdots,\lambda_n(t)]$. 试证:若$H(t)$有谱形式$H(t)=\sum_{j=1}^{n}\lambda_j(t)E_j(t)$,则对任意实变量的实值函数$f$,只要它在所有$\lambda_i(t)$的某邻域内连续可微,便有求导的 **Daleckii-Krein 公式**:

$$\frac{\mathrm{d}}{\mathrm{d}t}f(H(t))=\sum_{i,j=1}^{n}\frac{f(\lambda_i(t))-f(\lambda_j(t))}{\lambda_i(t)-\lambda_j(t)}E_i(t)\frac{\mathrm{d}}{\mathrm{d}t}H(t)E_j(t),$$

式中，若 $\lambda_i(t)=\lambda_j(t)$，则用 $f'(\lambda_i(t))$ 代替相应的差商.

5. 设 $A\in \boldsymbol{M}_n(\mathbb{C})$ 满足 $\|A\|_2<1$. 下面的复变量 λ 的矩阵值函数：

$$T_A(\lambda)=(I-AA^*)^{-1/2}(\lambda I+A)(I+\lambda A^*)^{-1}(I-A^*A)^{1/2}$$

叫作关于 A 的**特征函数**. 试证：$T_A(\cdot)$ 将复平面 $\mathbb{C}$ 内单位圆周 $\partial\Delta=\{\lambda\in\mathbb{C}:|\lambda|=1\}$ 映射到 $\mathscr{U}_n$（n 阶酉矩阵集合）内. $T_A(0)=A$. 并且，对任意 $\lambda\in\partial\Delta$，有连续矩阵值函数 $B(t)$，使得 $B(0)=I$，$B(1)=T_A(\lambda)$，且 $B(t)\in\mathscr{U}_n$，$\forall t\in[0,1]$.

6. 试证：若 $A\in\boldsymbol{M}_n(\mathbb{C})$ 满足 $\|A\|_2<1$，则 A 属于 $\mathscr{U}_n$ 的闭凸包.（提示：应用题 5 结果与全纯矩阵值函数的均值定理）

7. 试应用(10)式证明公式(7).

3.3.2　矩阵函数的积分形式定义与有关性质

假定 $f(\lambda)$ 为复值函数，它在开集 $\Omega\subset\mathbb{C}$ 内解析. 令

$$M_\Omega=\{A\in\boldsymbol{M}_n(\mathbb{C}):\sigma(A)\subset\Omega\}. \tag{1}$$

则对任意 $A\in M_\Omega$，$f(A)$ 都有定义，因此，$f(A)$ 可看为自 M_Ω 到 $\boldsymbol{M}_n(\mathbb{C})$ 内的映射. 在本小节后面将证明，M_Ω 为赋范线性空间 $\boldsymbol{M}_n(\mathbb{C})$ 内的开集，进而我们可以讨论 $f(A)$ 在 M_Ω 内的连续性、可微性与可积性.

给定 $A\in M_\Omega$. 设 $\Gamma\subset\Omega$ 为可求长的有向周线使得 $\sigma(A)$ 位于 Γ 包围的区域内部. 通常 Γ 可以由几个简单的周线组成，譬如，当 $\sigma(A)=\{\lambda_1,\lambda_2,\cdots,\lambda_s\}$ 时，可取

$$\Gamma=\bigcup_{j=1}^{s}\Gamma_j=\bigcup_{j=1}^{s}\{\lambda\in\mathbb{C}:|\lambda-\lambda_j|=\varepsilon_j\}, \tag{2}$$

式中，$\varepsilon_j>0$ 充分小使得 $\Gamma\subset\Omega$，且各 Γ_j 彼此不交，其内部只有 $\sigma(A)$ 的一个点. 这时，按复分析理论，

$$\int_\Gamma\phi(\lambda)\mathrm{d}\lambda=\sum_j\int_{\Gamma_j}\phi(\lambda)\mathrm{d}\lambda,$$

且 Γ 可以合适地选取使得

$$\frac{1}{2\pi\mathrm{i}}\int_\Gamma\frac{\mathrm{d}\lambda}{\lambda-\xi}=\begin{cases}1, & 如果\ \xi\in\sigma(A),\\ 0, & 如果\ \xi\notin\Omega.\end{cases} \tag{3}$$

这时 Cauchy 积分公式

$$f^{(k)}(\xi)=\frac{k!}{2\pi\mathrm{i}}\int_\Gamma(\lambda-\xi)^{-k-1}f(\lambda)\mathrm{d}\lambda,\qquad k=0,1,\cdots \tag{4}$$

对每一个在 Ω 内解析的函数 f 与每一个 $\xi\in\sigma(A)$ 成立. 通常，为了方便，称满足条件(3)的 Γ 为**在 $\boldsymbol{\Omega}$ 内围绕 $\boldsymbol{\sigma(A)}$ 的周线**. 注意(4)式中积分与这样周线 Γ 的选取无关.

取 $k=0$，(4)式变为 $f(\xi)=\frac{1}{2\pi\mathrm{i}}\int_\Gamma(\lambda-\xi)^{-1}f(\lambda)\mathrm{d}\lambda$，$\forall\xi\in\sigma(A)$. 受它的启发自然要问：对于上述的复值函数 f 与 $A\in M_\Omega$，$f(A)$ 是否有诸如 $f(A)=\frac{1}{2\pi\mathrm{i}}\int_\Gamma f(\lambda)(\lambda I-A)^{-1}\mathrm{d}\lambda$ 的表达式呢？下面定理给出它的肯定的答案.

定理 1 设 f 为复值函数，在开集 $\Omega\subset\mathbb{C}$ 内解析，$A\in M_\Omega$，且 Γ 为在 Ω 内围绕 $\sigma(A)$ 的周线. 则有

$$f(A)=\frac{1}{2\pi\mathrm{i}}\int_\Gamma f(\lambda)(\lambda I-A)^{-1}\mathrm{d}\lambda. \tag{5}$$

证明 显然在题设条件下，$f(A)$ 有定义. 另一方面，由于 $\lambda\in\Gamma$ 故 $\lambda\notin\sigma(A)$，因而 $(\lambda I-A)^{-1}$ 为 Γ 上的连续矩阵值函数. 于是，(5)式右端积分有意义且为 $\boldsymbol{M}_n(\mathbb{C})$ 中的矩阵.

按照 $f(A)$ 的谱分解定理(3.2.2 小节定理 1)，我们只要证明下式即可：

$$\frac{1}{2\pi\mathrm{i}}\int_\Gamma f(\lambda)(\lambda I-A)^{-1}\mathrm{d}\lambda=\sum_{k=1}^{s}\sum_{j=0}^{m_k-1}f^{(j)}(\lambda_k)E_{kj}, \tag{6}$$

这里假定 A 的最小多项式 $m(\lambda)=(\lambda-\lambda_1)^{m_1}\cdots(\lambda-\lambda_s)^{m_s}$. 但由习题 3.2.2 题 1

$$R_\lambda=(\lambda I-A)^{-1}=\sum_{k=1}^{s}\sum_{j=0}^{m_k-1}\frac{j!}{(\lambda-\lambda_k)^{j+1}}E_{kj}. \tag{7}$$

将(7)式代入(6)式左端积分，按公式(4)得出，

$$\frac{1}{2\pi\mathrm{i}}\int_\Gamma f(\lambda)R_\lambda\mathrm{d}\lambda=\sum_{k=1}^{s}\sum_{j=0}^{m_k-1}\frac{j!}{2\pi\mathrm{i}}\int_\Gamma\frac{f(\lambda)}{(\lambda-\lambda_k)^{j+1}}\mathrm{d}\lambda E_{kj}=\sum_{k=1}^{s}\sum_{j=0}^{m_k-1}f^{(j)}(\lambda_k)E_{kj}.$$

上式中间式子诸积分的被积函数为 $\Omega\backslash\sigma(A)$ 内的解析函数，按 Cauchy 定理，这些积分因而(5)式内积分均不依赖于 Γ 的选择(只要 Γ 为在 Ω 内围绕 $\sigma(A)$ 的周线). □

我们指出，定理 1 中(5)式有时也作为对于解析复值函数 $f(\lambda)$ 的 $f(A)$ 的定义. 它类似于复分析中 Cauchy 积分公式. 定理 1 表明，它与原先的 $f(A)$ 的两种定义相容.

取 $f(\lambda)=1$ 或 $f(\lambda)=\lambda^k$，$k=\pm1,\pm2,\cdots$，定理 1 直接蕴涵

$$\begin{aligned}I&=\frac{1}{2\pi\mathrm{i}}\int_\Gamma(\lambda I-A)^{-1}\mathrm{d}\lambda,\\A^k&=\frac{1}{2\pi\mathrm{i}}\int_\Gamma\lambda^k(\lambda I-A)^{-1}\mathrm{d}\lambda,\qquad k=\pm1,\pm2,\cdots,\end{aligned} \tag{8}$$

式中，各记号涵义同定理 1，但当 $k<0$ 时，应假定 $0\notin\sigma(A)$，即 $A\in\boldsymbol{M}_n(\mathbb{C})$ 非奇异.

有了定理 1，还可以将 $A\in\boldsymbol{M}_n(\mathbb{C})$ 的分量 E_{kj} $(0\leqslant j\leqslant m_k-1,1\leqslant k\leqslant s)$ 用积分形式表示. 事实上，考虑 $f(\lambda)=(\lambda-\lambda_p)^r$ $(0\leqslant r\leqslant m_p-1,1\leqslant p\leqslant s)$，且取 Γ 与 Γ_p 如(2)式. 应用(7)式，

$$\int_{\Gamma_p}(\lambda-\lambda_p)^r(\lambda I-A)^{-1}\mathrm{d}\lambda=\sum_{k=1}^{s}\sum_{j=0}^{m_k-1}j!E_{kj}\int_{\Gamma_p}\frac{(\lambda-\lambda_p)^r}{(\lambda-\lambda_k)^{j+1}}\mathrm{d}\lambda.$$

上式右端被积函数 $(\lambda-\lambda_p)^r(\lambda-\lambda_k)^{-j-1}$ 当 $k\neq p$ 或 $k=p$ 但 $r-j-1\geqslant0$ 时在 Γ_p 围

绕的区域内解析,因而$\int_{\Gamma_p}(\lambda-\lambda_p)^r(\lambda-\lambda_k)^{-j-1}\mathrm{d}\lambda=0$;当$k=p$但$r-j-1<0$时,按公式(4)(取$f(\lambda)\equiv 1$),

$$\int_{\Gamma_p}(\lambda-\lambda_p)^{r-j-1}\mathrm{d}\lambda=\begin{cases}2\pi\mathrm{i}, & \text{如果 } j=r,\\ 0, & \text{如果 } j>r.\end{cases}$$

因此,$\int_{\Gamma_p}(\lambda-\lambda_p)^r(\lambda I-A)^{-1}\mathrm{d}\lambda=r!(2\pi\mathrm{i})E_{pr}(0\leqslant r\leqslant m_p-1,1\leqslant p\leqslant s)$.

综上所述,我们证明了下面结果.

定理 2　设$A\in \boldsymbol{M}_n(\mathbb{C})$有最小多项式$m(\lambda)=(\lambda-\lambda_1)^{m_1}\cdots(\lambda-\lambda_s)^{m_s}$,$\Gamma$与$\Gamma_j$如(2)式.则对$j=0,1,\cdots,m_k-1;k=1,\cdots,s$,

$$E_{kj}=\frac{1}{j!2\pi\mathrm{i}}\int_{\Gamma_k}(\lambda-\lambda_k)^j(\lambda I-A)^{-1}\mathrm{d}\lambda,\tag{9}$$

特别地,

$$E_{k0}=\frac{1}{2\pi\mathrm{i}}\int_{\Gamma_k}(\lambda I-A)^{-1}\mathrm{d}\lambda,\qquad 1\leqslant k\leqslant s.\tag{10}$$

现在讨论这里定义的矩阵函数$f(A)$的有关性质.显然它保持了 3.1 节与 3.2 节论及的性质.我们感兴趣的是$f(A)$作为变量$A\in M_\Omega$的矩阵函数的连续性与可微性.

首先证明由(1)式确定的M_Ω为赋范线性空间$\boldsymbol{M}_n(\mathbb{C})$(取$\|\cdot\|$为某算子范数)内的开集,即对任意$A\in M_\Omega$,存在正数$\delta$使得对满足$\|B\|<\delta$的所有$B\in \boldsymbol{M}_n(\mathbb{C})$都有$\sigma(A+B)\subset\Omega$.这是因为$\|(\lambda I-A)^{-1}\|$在$A$的预解集$\mathscr{R}(A)$内为$\lambda$的连续函数,且$\lim\limits_{\lambda\to\infty}\|(\lambda I-A)^{-1}\|=0$(因为$R_\infty=O$,见 3.3.1 小节定理 1),故存在正数$q$使得对所有$\lambda\in\mathbb{C}\backslash\Omega$,$\|(\lambda I-A)^{-1}\|<q$.现若$B\in\boldsymbol{M}_n(\mathbb{C})$满足$\|B\|<1/q$,且$\lambda\in\mathbb{C}\backslash\Omega$,则$\|(\lambda I-A)^{-1}B\|\leqslant\|(\lambda I-A)^{-1}\|\,\|B\|<1$,从而

$$\lambda I-(A+B)=(\lambda I-A)(I-(\lambda I-A)^{-1}B)\tag{11}$$

为$\boldsymbol{M}_n(\mathbb{C})$内非奇异矩阵,这表明$\lambda\notin\sigma(A+B)$.因此,$\sigma(A+B)\subset\Omega$.

接着讨论矩阵函数$f(A)$作为M_Ω到$\boldsymbol{M}_n(\mathbb{C})$内映射的可微性,所用的$\boldsymbol{M}_n(\mathbb{C})$或$\mathscr{L}(\mathscr{S})$上的范数均为算子范数(见第二章 2.2.2 小节(21)式).

定义 3　设$\mathscr{S}$为有限维赋范空间,Ω为$\mathscr{S}$内开集,F映射Ω到$\mathscr{S}$内.对$\boldsymbol{a}\in\Omega$,若存在$\Lambda\in\mathscr{L}(\mathscr{S})$($\mathscr{S}$上有界线性算子的集合)使得

$$\lim_{\boldsymbol{x}\to 0}\frac{\|F(\boldsymbol{a}+\boldsymbol{x})-F(\boldsymbol{a})-\Lambda\boldsymbol{x}\|}{\|\boldsymbol{x}\|}=0,\tag{12}$$

则称Λ为F在$\boldsymbol{a}\in\Omega$处的 **Fréchet 导数**,记为$(DF)_a$.假如对每一个$\boldsymbol{a}\in\Omega$,$(DF)_a$存在,则称F为Ω上的**全纯函数**.若$\boldsymbol{a}\to(DF)_a$还是自Ω到$\mathscr{L}(\mathscr{S})$内的连续映射,则称F在Ω内为**连续可微的**.

选取$\boldsymbol{M}_n(\mathbb{C})$为定义 3 中$\mathscr{S}$,$M_\Omega$为$\Omega$,$f(A)$定义的自$M_\Omega$到$\boldsymbol{M}_n(\mathbb{C})$内映射为$F$,

下面定理将证明,$f:M_\Omega \to \boldsymbol{M}_n(\mathbb{C})$为连续可微的映射,只要 $f(\lambda)$在 Ω 内解析.

定理 4 设 Ω 为$\mathbb{C}$内开集,f 为复值函数,它在 Ω 内解析,且 $A\in M_\Omega$. 则由(5)式确定的映射 $f:M_\Omega\to \boldsymbol{M}_n(\mathbb{C})$在 M_Ω 内连续可微,且对任意 $X\in \boldsymbol{M}_n(\mathbb{C})$,

$$(Df)_A(X)=\frac{1}{2\pi\mathrm{i}}\int_\Gamma f(\lambda)(\lambda I-A)^{-1}X(\lambda I-A)^{-1}\mathrm{d}\lambda, \tag{13}$$

式中,Γ 为在 Ω 内围绕 $\sigma(A)$的任意周线.

证明 对给定的 $A\in \boldsymbol{M}_n(\mathbb{C})$,$(\lambda I-A)^{-1}$为 A 的预解集 $\mathscr{R}(A)$内的全纯矩阵值函数,因而有正数 $q>0$ 使得 $\|(\lambda I-A)^{-1}\|<q$, $\forall\lambda\in\Gamma$. 现取 $Y\in \boldsymbol{M}_n(\mathbb{C})$满足 $\|Y\|<1/(2q)$. 则对任意 $\lambda\in\Gamma$,$\lambda I-A-Y$ 可逆. 这时,由(5)式与 3.3.1 小节(6)式(取 $A(t)=tI-A$ 与 $A(t+h)=tI-A-Y$),

$$\begin{aligned}&\left\| f(A+Y)-f(A)-\frac{1}{2\pi\mathrm{i}}\int_\Gamma f(\lambda)(\lambda I-A)^{-1}Y(\lambda I-A)^{-1}\mathrm{d}\lambda\right\|\\
=&\left\|\frac{1}{2\pi\mathrm{i}}\int_\Gamma f(\lambda)[(\lambda I-A-Y)^{-1}-(\lambda I-A)^{-1}-(\lambda I-A)^{-1}Y(\lambda I-A)^{-1}]\mathrm{d}\lambda\right\|\\
\leqslant&\max_{\lambda\in\Gamma}|f(\lambda)|\cdot 2q^3\|Y\|^2\cdot(\Gamma\text{的长度})/(2\pi).\end{aligned}$$

因此,按定义 3,$(Df)_A(Y)=\dfrac{1}{2\pi\mathrm{i}}\displaystyle\int_\Gamma f(\lambda)(\lambda I-A)^{-1}Y(\lambda I-A)^{-1}\mathrm{d}\lambda$. 此外,若 $A_p\in M_\Omega$, $p=1,2,\cdots$,且 $A_p\to A\in M_\Omega$,则(13)式中的 Γ 也是在 Ω 内围绕 $\sigma(A_p)$的周线(但可能要去掉有限个 p). 按刚才证明,$(Df)_{A_p}$ 有定义,且由(13)式给出(只要用 A_p 代替 A 即可). 根据 3.3.1 小节(6)式可推出,

$$\begin{aligned}&\|(\lambda I-A_p)^{-1}-(\lambda I-A)^{-1}\|\leqslant\|(\lambda I-A)^{-1}(A_p-A)(\lambda I-A)^{-1}\|\\
&+2\|(\lambda I-A)^{-1}\|^3\|A_p-A\|^2\\
&\leqslant\|(\lambda I-A)^{-1}\|^2\|A_p-A\|+2\|(\lambda I-A)^{-1}\|^3\|A_p-A\|^2,\end{aligned}$$

于是,

$$(\lambda I-A_p)^{-1}\to(\lambda I-A)^{-1},\qquad p\to\infty,$$

在 Γ 上一致成立. 因此,对任意给定的 $Y\in \boldsymbol{M}_n(\mathbb{C})$, $\|(Df)_{A_p}(Y)-(Df)_A(Y)\|\to 0(p\to\infty)$,此蕴涵$(Df)_{A_p}\to(Df)_A(p\to\infty)$,即映射 $A\to(Df)_A$为 M_Ω 内的连续映射. □

习题 3.3.2

1. 应用定理 2 中的记号试证:

$$\lambda_k E_{k0}+E_{k1}=AE_{k0}=\frac{1}{2\pi\mathrm{i}}\int_{\Gamma_k}\lambda R_\lambda\mathrm{d}\lambda.$$

2. 设 $J\in \boldsymbol{M}_n(\mathbb{C})$为某 n 阶矩阵的 Jordan 正规形式. 应用定理 2 证明:J 的分量 $E_{k0}(1\leqslant k\leqslant s)$有形式 $\mathrm{diag}[d_{k1},d_{k2},\cdots,d_{kn}]$,式中各 d_{kj} 等于 0 或 1, $1\leqslant k\leqslant s$, $1\leqslant j\leqslant n$.

3. 试证:若 $A\in \boldsymbol{M}_n(\mathbb{C})$所有特征值有负的实部,则

$$A^{-1}=-\int_0^{\infty}e^{At}\,dt,$$

并且一般地，如果 A 的所有特征值在半平面 $\{\lambda\in\mathbb{C}:\mathrm{Re}\lambda<\mathrm{Re}\lambda_0\}$ 内，则

$$(\lambda_0 I-A)^{-1}=\int_0^{\infty}e^{-(\lambda_0 I-A)t}\,dt.$$

4. 应用公式(8)证明第二章 2.2.2 小节定理 11.

3.4 对微分方程的应用

3.4.1 一阶常系数常微分方程组解的表达式

考虑一阶常微分方程组

$$\begin{aligned}\dot{x}_1(t)&=a_{11}(t)x_1(t)+a_{12}(t)x_2(t)+\cdots+a_{1n}(t)x_n(t),\\ \dot{x}_2(t)&=a_{21}(t)x_1(t)+a_{22}(t)x_2(t)+\cdots+a_{2n}(t)x_n(t),\\ &\ \ \vdots\\ \dot{x}_n(t)&=a_{n1}(t)x_1(t)+a_{n2}(t)x_2(t)+\cdots+a_{nn}(t)x_n(t),\end{aligned}$$

这里，$a_{ij}(t)(1\leqslant i,j\leqslant n)$ 为给定的函数. 记 $\boldsymbol{x}(t)=(x_1(t),\cdots,x_n(t))^{\mathrm{T}}$，$\dot{\boldsymbol{x}}(t)=(\dot{x}_1(t),\cdots,\dot{x}_n(t))^{\mathrm{T}}$，$A(t)=(a_{ij}(t))(1\leqslant i,j\leqslant n)$. 上述微分方程组可写成矩阵-向量形式：

$$\dot{\boldsymbol{x}}(t)=A(t)\boldsymbol{x}(t). \tag{1}$$

这是齐次形式. 一般地还需要考虑非齐次一阶常微分方程组

$$\dot{\boldsymbol{x}}(t)-A(t)\boldsymbol{x}(t)=\boldsymbol{f}(t), \tag{2}$$

式中，$\boldsymbol{f}(t)=(f_1(t),\cdots,f_n(t))^{\mathrm{T}}$ 为已知的连续或逐段连续的向量值函数.

在本小节我们讨论常微分方程组(1)解的存在唯一性问题、解空间的结构，以及在常系数情形微分方程组(1)与(2)解的表达式.

定理 1　若 $A(t)$元素 $a_{ij}(t)(i,j=1,2,\cdots,n)$为 $t\geqslant 0$ 的连续函数，且 $\boldsymbol{c}=(c_1,\cdots,c_n)^{\mathrm{T}}\in\mathbb{C}^n$ 为常数向量，则对 $t\geqslant 0$ 存在唯一的解 $\boldsymbol{x}(t)$满足微分方程组(1)与初值条件 $\boldsymbol{x}(0)=\boldsymbol{c}$. 并且，此解可表示为

$$\boldsymbol{x}(t)=X(t)\boldsymbol{c}, \tag{3}$$

式中 $X(t)$为下列矩阵常微分方程初值问题的解：

$$\frac{\boldsymbol{d}}{\boldsymbol{dt}}X(t)=A(t)X(t),\qquad X(0)=I. \tag{4}$$

证明　考虑初值问题(4)的积分方程等价形式：

$$X(t)=I+\int_0^t A(s)X(s)\mathrm{d}s. \tag{5}$$

取 $X_0(t)=I$ 与 $X_{p+1}(t)=I+\int_0^t A(s)X_p(s)\mathrm{d}s$，$p=0,1,\cdots$. 则有

$$X_{p+1}(t)-X_p(t)=\int_0^t A(s)(X_p(s)-X_{p-1}(s))\mathrm{d}s,\qquad p=1,2,\cdots,$$

$$X_1(t)-X_0(t)=\int_0^t A(s)\,\mathrm{d}s.$$

设 $\|\cdot\|$ 为 $\mathbb{C}^n$ 上向量范数或它导出的矩阵范数. 取定 $t_1>0$, 令 $M=\max\limits_{0\leqslant t\leqslant t_1}\|A(t)\|$, 按题设, $M<\infty$. 这时由上式推出,

$$\begin{aligned}\|X_{p+1}(t)-X_p(t)\| &\leqslant \int_0^t \|A(s)\|\ \|X_p(s)-X_{p-1}(s)\|\,\mathrm{d}s\\ &\leqslant M\int_0^t \|X_p(s)-X_{p-1}(s)\|\,\mathrm{d}s, \qquad 0\leqslant t\leqslant t_1,\end{aligned}$$

$$\|X_1(t)-X_0(t)\|\leqslant\int_0^t\|A(s)\|\,\mathrm{d}s\leqslant Mt, \qquad 0\leqslant t\leqslant t_1.$$

于是,

$$\|X_{p+1}(t)-X_p(t)\|\leqslant\frac{M^{p+1}}{(p+1)!}t^{p+1}, \qquad 0\leqslant t\leqslant t_1.$$

由此看出, $\sum\limits_{p=0}^{\infty}(X_{p+1}(t)-X_p(t))$ 对 $0\leqslant t\leqslant t_1$ 一致收敛, 因而当 $p\to\infty$ 时, $X_p(t)$ 在 $0\leqslant t\leqslant t_1$ 上一致收敛于某个 n 阶矩阵值函数 $X(t)$, 它是积分方程(5)因而是初值问题(4)的解. 按题设, $A(t)$ 对 $t\geqslant 0$ 为连续的, 因此 t_1 可取任意大, 这样我们对 $t\geqslant 0$ 得到初值问题(4)的一个解. 容易验明, $\boldsymbol{x}(t)=X(t)\boldsymbol{c}$ 为微分方程组(1)带有初值条件 $\boldsymbol{x}(0)=\boldsymbol{c}$ 的解.

下面证明初值问题(4)解的唯一性. 设 $Y(t)$ 为问题(4)的另一个解, 则有

$$X(t)-Y(t)=\int_0^t A(s)(X(s)-Y(s))\,\mathrm{d}s. \tag{6}$$

由于 $Y(t)$ 在 $t\geqslant 0$ 区间上可微, 故它连续. 令 $M_1=\max\limits_{0\leqslant t\leqslant t_1}\|X(t)-Y(t)\|$. (6)式蕴涵

$$\|X(t)-Y(t)\|\leqslant M_1\int_0^t\|A(s)\|\,\mathrm{d}s, \qquad 0\leqslant t\leqslant t_1.$$

因此,

$$\begin{aligned}\|X(t)-Y(t)\| &\leqslant \int_0^t\|A(s)\|\ \|X(s)-Y(s)\|\,\mathrm{d}s\\ &\leqslant M_1\int_0^t\|A(s)\|\left[\int_0^s\|A(s_1)\|\,\mathrm{d}s_1\right]\mathrm{d}s\\ &\leqslant M_1\left[\int_0^t\|A(s)\|\,\mathrm{d}s\right]^2\Big/2, \qquad 0\leqslant t\leqslant t_1.\end{aligned}$$

按此推行下去, 可得

$$\|X(t)-Y(t)\|\leqslant M_1\left[\int_0^t\|A(s)\|\,\mathrm{d}s\right]^{p+1}\Big/(p+1)!, \qquad 0\leqslant t\leqslant t_1.$$

令 $p\to\infty$ 便有 $\|X(t)-Y(t)\|\leqslant 0$, 因而 $X(t)=Y(t)$, $0\leqslant t\leqslant t_1$. 由于 t_1 可取任意大, 故有 $X(t)=Y(t)$, $\forall\, t\geqslant 0$.

按照上述类似的方法也可证明原初值问题 $\dot{\boldsymbol{x}}(t)=A(t)\boldsymbol{x}(t),\boldsymbol{x}(0)=\boldsymbol{c}$ 的解的唯一性. □

若 $\boldsymbol{x}(t)$ 为微分方程组(1)的解,算出 $\boldsymbol{x}(0)$,则从定理 1 看出,$\boldsymbol{x}(t)=X(t)\boldsymbol{x}(0)$,这里,$X(t)$ 为初值问题(4)的唯一解. 因此再按定理 1 有如下结论.

定理 2　微分方程组(1)的解 $\boldsymbol{x}(t)$ 可表达为 $\boldsymbol{x}(t)=X(t)\boldsymbol{z}$ 的形式,这里,$\boldsymbol{z}\in\mathbb{C}^n$,$X(t)$ 为初值问题(4)的唯一解.

我们指出,定理 2 表明微分方程组(1)的任一个解 $\boldsymbol{x}(t)$ 都是 $X(t)$ 的 n 个列向量的线性组合. 显然,这些列向量均为微分方程组(1)的解. 下一结果表明 $\det X(t)\neq 0,\forall t\geqslant 0$,因而微分方程组(1)的解空间 $\mathscr{S}$ 的维数 $\dim\mathscr{S}=n$,且 $X(t)$ 的 n 个列向量组成 $\mathscr{S}$ 的基.

定理 3　设 $A(t)=(a_{ij}(t))(1\leqslant i,j\leqslant n)$ 在区间 $t\geqslant 0$ 上连续. 则初值问题(4)的解 $X(t)$ 在区间 $t\geqslant 0$ 上非奇异.

证明　为了证明本定理结果,只需证明下列的所谓 **Jacobi 恒等式**:

$$\det X(t)=\exp\int_0^t \operatorname{tr}(A(s))\,\mathrm{d}s,\qquad \forall t\geqslant 0. \tag{7}$$

下面仅对二阶情形验证恒等式(7),一般情形可类似地验证. 令 $X(t)=(x_{ij}(t))(1\leqslant i,j\leqslant 2)$. 则初值问题(4)变为

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix}x_{11}(t) & x_{12}(t)\\ x_{21}(t) & x_{22}(t)\end{bmatrix}=\begin{bmatrix}a_{11}(t) & a_{12}(t)\\ a_{21}(t) & a_{22}(t)\end{bmatrix}\begin{bmatrix}x_{11}(t) & x_{12}(t)\\ x_{21}(t) & x_{22}(t)\end{bmatrix},$$

$$\begin{bmatrix}x_{11}(0) & x_{12}(0)\\ x_{21}(0) & x_{22}(0)\end{bmatrix}=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}.$$

因此,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}(\det X(t)) &= \det\begin{bmatrix}\dot{x}_{11}(t) & \dot{x}_{12}(t)\\ x_{21}(t) & x_{22}(t)\end{bmatrix}+\det\begin{bmatrix}x_{11}(t) & x_{12}(t)\\ \dot{x}_{21}(t) & \dot{x}_{22}(t)\end{bmatrix}\\ &= \det\begin{bmatrix}a_{11}(t)x_{11}(t)+a_{12}(t)x_{21}(t) & a_{11}(t)x_{12}(t)+a_{12}(t)x_{22}(t)\\ x_{21}(t) & x_{22}(t)\end{bmatrix}\\ &\quad+\det\begin{bmatrix}x_{11}(t) & x_{12}(t)\\ a_{21}(t)x_{11}(t)+a_{22}(t)x_{21}(t) & a_{21}(t)x_{12}(t)+a_{22}(t)x_{22}(t)\end{bmatrix}\\ &= a_{11}(t)\det\begin{bmatrix}x_{11}(t) & x_{12}(t)\\ x_{21}(t) & x_{22}(t)\end{bmatrix}+a_{22}(t)\det\begin{bmatrix}x_{11}(t) & x_{12}(t)\\ x_{21}(t) & x_{22}(t)\end{bmatrix}\\ &= \operatorname{tr}A(t)\cdot\det X(t).\end{aligned}$$

求解之得到

$$\det X(t)=\exp\int_0^t \operatorname{tr}A(s)\,\mathrm{d}s,\qquad \forall t\geqslant 0,$$

因为 $\det X(0)=\det I=1$. □

现在分别地研究一阶常系数微分方程组

$$\dot{\boldsymbol{x}}(t)=A\boldsymbol{x}(t), \tag{8}$$

与

$$\dot{\boldsymbol{x}}(t)=A\boldsymbol{x}(t)+\boldsymbol{f}(t) \tag{9}$$

的解的表达式,式中,$A=(a_{ij})\in\boldsymbol{M}_n(\mathbb{C})$与$t$无关,其他记号涵义同(1)与(2)式.

由定理1知道,对任意给定的向量$\boldsymbol{x}_0\in\mathbb{C}^n$,初值问题:

$$\dot{\boldsymbol{x}}(t)=A\boldsymbol{x}(t),\qquad \dot{\boldsymbol{x}}(0)=\boldsymbol{x}_0 \tag{10}$$

有唯一解,下面定理指出它的唯一解的形式.

定理4　初值问题(10)的唯一解为$\boldsymbol{x}(t)=\mathrm{e}^{At}\boldsymbol{x}_0$.

证明　显然只要证明:$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{At}=A\mathrm{e}^{At}$,但此已由习题3.3.1题1证实.　□

按照谱分解定理,我们有

$$\mathrm{e}^{At}=\sum_{k=1}^{s}\sum_{j=0}^{m_k-1}t^j\mathrm{e}^{\lambda_k}E_{kj},$$

因此,初值问题(10)的唯一解还可写成

$$\boldsymbol{x}(t)=\sum_{k=1}^{s}\sum_{j=0}^{m_k-1}t^j\mathrm{e}^{\lambda_k}E_{kj}\boldsymbol{x}_0=\sum_{k=1}^{s}(\boldsymbol{z}_{k0}+\boldsymbol{z}_{k1}t+\cdots+\boldsymbol{z}_{k,m_k-1}t^{m_k-1})\mathrm{e}^{\lambda_k t}, \tag{11}$$

式中,$\boldsymbol{z}_{kj}=E_{kj}\boldsymbol{x}_0,j=0,1,\cdots,m_k-1;k=1,2,\cdots,s$. 特殊地,若$A\in\boldsymbol{M}_n(\mathbb{C})$为简单矩阵,则上式退化为

$$\boldsymbol{x}(t)=\sum_{k=1}^{s}\boldsymbol{z}_{k0}\mathrm{e}^{\lambda_k t}. \tag{12}$$

对于非齐次方程组(9)的初值问题:

$$\dot{\boldsymbol{x}}(t)=A\boldsymbol{x}(t)+\boldsymbol{f}(t),\qquad \boldsymbol{x}(0)=\boldsymbol{x}_0, \tag{13}$$

这里,假定$\boldsymbol{f}(t)=(f_1(t),f_2(t),\cdots,f_n(t))^{\mathrm{T}}$为区间$t\geqslant0$上的逐段连续向量值函数,我们有

定理5　初值问题(13)的解可以唯一地表示为

$$\boldsymbol{x}(t)=\mathrm{e}^{At}\boldsymbol{x}_0+\int_0^t\mathrm{e}^{A(t-\tau)}\boldsymbol{f}(\tau)\mathrm{d}\tau. \tag{14}$$

证明　问题(13)解的唯一性为定理1的直接推论. 根据事实:$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{At}=A\mathrm{e}^{At}$,若$\boldsymbol{x}(t)$为初值问题(13)的解,则

$$\mathrm{e}^{-At}(\dot{\boldsymbol{x}}(t)-A\boldsymbol{x}(t))=\frac{\mathrm{d}}{\mathrm{d}t}(\mathrm{e}^{-At}\boldsymbol{x}(t))=\mathrm{e}^{-At}\boldsymbol{f}(t).$$

因此,

$$\mathrm{e}^{-At}\boldsymbol{x}(t)=\boldsymbol{x}_0+\int_0^t\mathrm{e}^{-A\tau}\boldsymbol{f}(\tau)\mathrm{d}\tau,$$

从而有

$$x(t)=e^{At}x_0+\int_0^t e^{A(t-\tau)}f(\tau)d\tau,$$

直接代入验证便知，它为问题(13)的解. □

例 6　求解常微分方程组初值问题：

$$\dot{x}(t)=Ax(t),\qquad x(0)=x_0=(1,-1,0)^{\mathrm{T}}$$

与

$$\dot{x}(t)=Ax(t)+f(t),\qquad x(0)=x_0=(1,-1,0)^{\mathrm{T}},$$

其中，

$$A=\begin{bmatrix}3&-1&1\\2&0&1\\1&-1&2\end{bmatrix}\quad 与\quad f(t)=\begin{bmatrix}0\\0\\e^{2t}\end{bmatrix}=e^{2t}e_3$$

在 3.2.2 小节例 6 中已求得 e^{At}. 按定理 4，本例中齐次初值问题的解为

$$\begin{aligned}x(t)&=\begin{bmatrix}(t+1)e^{2t}&-te^{2t}&te^{2t}\\-e^t+(t+1)e^{2t}&e^t-te^{2t}&te^{2t}\\-e^t+e^{2t}&e^t-e^{2t}&e^{2t}\end{bmatrix}\begin{bmatrix}1\\-1\\0\end{bmatrix}\\&=((2t+1)e^{2t},-2e^t+(2t+1)e^{2t},-2e^t+2e^{2t})^{\mathrm{T}}\\&=e^{2t}(2t+1,2t+1,2)^{\mathrm{T}}+e^t(0,-2,-2)^{\mathrm{T}}.\end{aligned}\tag{15}$$

对于非齐次初值问题，按定理 5，我们只要求积分

$$y(t)=\int_0^t e^{A(t-\tau)}f(\tau)d\tau=\sum_{k=1}^{s}\sum_{j=0}^{m_k-1}E_{kj}\int_0^t(t-\tau)^j e^{\lambda_k(t-\tau)}f(\tau)d\tau$$

即可. 但从 3.2.2 小节例 6 知道，$\lambda_1=1,m_1=1;\lambda_2=2,m_2=2$ 以及 E_{10},E_{20} 与 E_{21} 的具体形式，故上式右端为

$$\begin{aligned}&E_{10}\int_0^t e^{t-\tau}(0,0,e^{2\tau})^{\mathrm{T}}d\tau+E_{20}\int_0^t e^{2(t-\tau)}(0,0,e^{2\tau})^{\mathrm{T}}d\tau\\&\quad+E_{21}\int_0^t(t-\tau)e^{2(t-\tau)}(0,0,e^{2\tau})^{\mathrm{T}}d\tau\\&=e^t(e^t-1)E_{10}e_3+te^{2t}E_{20}e_3+\frac{1}{2}t^2e^{2t}E_{21}e_3\\&=\left(\frac{1}{2}t^2e^{2t},\frac{1}{2}t^2e^{2t},te^{2t}\right)^{\mathrm{T}}.\end{aligned}$$

因此，应用(15)式，本例中非齐次初值问题解为

$$\begin{aligned}x(t)&=e^{At}x_0+y(t)\\&=\left(\left(\frac{1}{2}t^2+2t+1\right)e^{2t},\left(\frac{1}{2}t^2+2t+1\right)e^{2t}-2e^t,(t+2)e^{2t}-2e^t\right)^{\mathrm{T}}.\end{aligned}$$

□

最后，我们指出当 $A\in \boldsymbol{M}_n(\mathbb{C})$ 为一般矩阵，其 s 个不同特征值 $\lambda_1,\lambda_2,\cdots,\lambda_s$ 分别有指标 $m_1,m_2,\cdots,m_s$ 时，由(11)式看出，微分方程组(8)的解空间 $\mathcal{S}$(维数为 n)有基

$$\left\{\sum_{k=1}^{s}\sum_{j=0}^{m_k-1}t^j e^{\lambda_k t}E_{kj}\boldsymbol{e}_l\right\}_{l=1}^{n}=\{e^{At}\boldsymbol{e}_l\}_{l=1}^{n},\tag{16}$$

这里 $\boldsymbol{e}_l$ 为$\mathbb{C}^n$ 的第 l 个单位坐标向量，换句话说，非奇异矩阵 e^{At} 的 n 个列向量组成 $\mathcal{S}$的基.

习题 3.4.1

1. 设 $A\in \boldsymbol{M}_3(\mathbb{C})$ 为简单矩阵，它有如下形式：

$$A=PDP^{-1}=\begin{bmatrix}1&0&1\\1&1&0\\0&1&1\end{bmatrix}\begin{bmatrix}2&0&0\\0&1&0\\0&0&1\end{bmatrix}\begin{bmatrix}\frac{1}{2}&\frac{1}{2}&-\frac{1}{2}\\-\frac{1}{2}&\frac{1}{2}&\frac{1}{2}\\\frac{1}{2}&-\frac{1}{2}&\frac{1}{2}\end{bmatrix}.$$

试求微分方程组(8)与(9)带有初值条件 $\boldsymbol{x}(0)=(-2,1,-1)^{\mathrm{T}}$ 的解，其中，$\boldsymbol{f}(t)=(0,e^t,e^{-t})^{\mathrm{T}}$.

2. 设 $A\in \boldsymbol{M}_3(\mathbb{C})$ 有如习题 3.2.2 题 3 的形式. 应用那里求得的结果，求微分方程组(8)与(9)带有初始条件 $\boldsymbol{x}(0)=(0,1,0)^{\mathrm{T}}$ 与 $\boldsymbol{f}(t)=(0,e^t,0)^{\mathrm{T}}$ 的解.

3. 设 $A\in \boldsymbol{M}_n(\mathbb{C})$ 非奇异，$\boldsymbol{\beta}=A^{-1}\boldsymbol{b}$，其中 $\boldsymbol{b}\in\mathbb{C}^n$. 试证：若 $\boldsymbol{x}(t)$ 为 $\dot{\boldsymbol{x}}(t)=A\boldsymbol{x}(t)-\boldsymbol{b}$ 的解，则

$$\boldsymbol{x}(t)-\boldsymbol{\beta}=e^{At}(\boldsymbol{x}(0)-\boldsymbol{\beta}).$$

(这里，β 为该微分方程的**稳态解**)

4. 设 $A\in \boldsymbol{M}_n(\mathbb{C})$，且 $\boldsymbol{f}_0,\boldsymbol{f}_1,\cdots$为$\mathbb{C}^n$ 内给定的向量序列. 试证：差分方程

$$\boldsymbol{x}_{r+1}=A\boldsymbol{x}_r+\boldsymbol{f}_r,\qquad r=0,1,2,\cdots$$

的任一解可以表达为

$$\boldsymbol{x}_0=\boldsymbol{z},\boldsymbol{x}_1=A\boldsymbol{z}+\boldsymbol{f}_0,\qquad \boldsymbol{x}_{j+1}=A^{j+1}\boldsymbol{z}+\sum_{k=0}^{j}A^{j-k}\boldsymbol{f}_k,\qquad j=1,2,\cdots$$

这里，$\boldsymbol{z}\in\mathbb{C}^n$.

5. 设 $X(t)$为可微的 n 阶矩阵值函数，$t\geqslant 0$，它满足

$$\dot{X}(t)=AX(t)+X(t)B,\qquad X(0)=C,$$

式中，A,B 与 C 为 $\boldsymbol{M}_n(\mathbb{C})$ 中给定的矩阵. 试证明该问题解的唯一性，并用 A,B 与 C 表达解 $X(t)$.

3.4.2　可观测与可控的定常线性系统

在控制系统理论中常遇到如下的所谓**受控定常(时不变)系统**：

$$\dot{\boldsymbol{x}}(t)=A\boldsymbol{x}(t)+B\boldsymbol{u}(t)\tag{1}$$

$$\boldsymbol{y}(t)=C\boldsymbol{x}(t),\tag{2}$$

这里，$A\in \boldsymbol{M}_n(\mathbb{C})$，$B\in \boldsymbol{M}_{n,m}(\mathbb{C})$，$C\in \boldsymbol{M}_{r,n}(\mathbb{C})$给定，它们都与 t 无关，且 $\boldsymbol{x}(t)\in\mathbb{C}^n$ 与

$\boldsymbol{u}(t) \in \mathbb{C}^m$, $\forall t \geqslant 0$. 通常，$\boldsymbol{x}(t)$称为系统在时刻 t 的**状态向量**(state vector)；$\boldsymbol{u}(t)$称为系统在时刻 t 的**输入**(input)或**控制**(control)，它使系统的状态达到某种预期的目的. 在后面分析中，我们限定 $\boldsymbol{u}(t)$的可取范围为 $\mathscr{U} = \{\boldsymbol{u}(t) \in \mathbb{C}^m : \boldsymbol{u}(t)$为逐段连续的向量值函数$\}$，对每一个 $t \geqslant 0$，$\boldsymbol{y}(t) \in \mathbb{C}^r$ 叫作系统的**观测**或**输出**向量(observed or output vector).

为了控制一个定常系统(1)～(2)，需要了解系统的状态 $\boldsymbol{x}(t)$. 但一般不能直接测量到它，必须通过矩阵 C"过滤"后得到的观测向量 $\boldsymbol{y}(t)$反过来判断 $\boldsymbol{x}(t)$. 因此，基本问题之一是：能否通过观测向量 $\boldsymbol{y}(t)$确定出系统的全部状态，这就是所谓的**可观测性问题**. 掌握了系统的状态后，还要控制它使其达到预期的目的，这便是另一个基本问题，即所谓**可控性问题**.

在后面讨论中将看到，前述系统的可观测性问题依赖于矩阵对(C,A)的特性，而它的可控性问题则依赖于矩阵对(A,B)的特性. 为此，我们首先讨论矩阵对的可观测性和可控性的概念与有关结论.

现设 $X \in \boldsymbol{M}_{r,p}(\mathbb{C})$与 $T \in \boldsymbol{M}_p(\mathbb{C})$. 若 $\mathrm{Ker}X \neq \{\boldsymbol{0}\}$，即 X 的列向量线性相关，则矩阵对(X,T)称为 **p 阶容许对**(admissible pair of order $\boldsymbol{p}$). 这里的 r 与 p 虽可选为任意正整数，但在实用上常遇到的是 $r \leqslant p$(尤其是 $r=1 \leqslant p$)的情形. 倘若 $r<p$，则按第一章 1.1.1 小节定理 6，$\mathrm{nul}X = p - \mathrm{rank}X > 0$，因而有 $\mathrm{Ker}X \neq \{\boldsymbol{0}\}$，这时$(X,T)$总为 p 阶容许对.

令 $\mathscr{K}_1 = \mathrm{Ker}X \neq \{\boldsymbol{0}\}$与 $\mathscr{K}_2 = \mathrm{Ker}X \cap \mathrm{Ker}(XT)$. 则显然有 $\mathscr{K}_2 \subset \mathscr{K}_1$，且 $\mathscr{K}_2$ 可改写为

$$\mathscr{K}_2 = \mathrm{Ker}\begin{bmatrix} X \\ XT \end{bmatrix}.$$

一般地，令 $\mathscr{K}_l = \mathrm{Ker}X \cap \mathrm{Ker}(XT) \cap \cdots \cap \mathrm{Ker}(XT^{l-1})$，$l=1,2,\cdots$. 显然，$\mathscr{K}_l$ 为$\mathbb{C}^p$ 的子空间，

$$\mathscr{K}_1 \supset \mathscr{K}_2 \supset \mathscr{K}_3 \supset \cdots, \tag{3}$$

并且，$\mathscr{K}_l$ 可改写为

$$\mathscr{K}_l = \mathrm{Ker}\begin{bmatrix} X \\ XT \\ XT^2 \\ \vdots \\ XT^{l-1} \end{bmatrix}, \quad l = 1,2,\cdots. \tag{4}$$

由于 $\mathscr{K}_l$ 为有限维的，故$\mathbb{C}^p$ 中子空间序列(3)显然不可能都是严格包含的，即存在正整数 j 使得 $\mathscr{K}_j = \mathscr{K}_{j+1}$，下面引理告诉我们，若 $\mathscr{K}_j = \mathscr{K}_{j+1}$，则 $\mathscr{K}_j = \mathscr{K}_l$，$\forall l > j$，这表明：(3)式中只有前面的有限多个$\mathbb{C}^p$ 的不同子空间.

引理 1 设(X,T)为 p 阶容许对. 若在序列(3)中 $\mathscr{K}_j = \mathscr{K}_{j+1}$，则 $\mathscr{K}_{j+1} = \mathscr{K}_{j+2} = \cdots$.

证明　考虑如下两个 $jr\times p$ 与 $(j+1)r\times p$ 的矩阵

$$\begin{bmatrix} X \\ XT \\ \vdots \\ XT^{j-1} \end{bmatrix} \quad 与 \quad \begin{bmatrix} X \\ XT \\ \vdots \\ XT^{j} \end{bmatrix}. \tag{5}$$

假如 $\mathscr{K}_j=\mathscr{K}_{j+1}$，则按第一章 1.1.1 小节定理 6，(5)式中两个矩阵具有相同的秩：$p-\dim\mathscr{K}_j$. 这意味着矩阵 XT^j 的每一行都是(5)式中头一个矩阵诸行向量的线性组合，因而存在一个 $r\times jr$ 矩阵 E 使得

$$XT^j = E\begin{bmatrix} X \\ XT \\ \vdots \\ XT^{j-1} \end{bmatrix}. \tag{6}$$

现设 $\boldsymbol{x}\in\mathscr{K}_{j+1}$，应用(6)式则有

$$XT^{j+1}\boldsymbol{x} = (XT^j)T\boldsymbol{x} = E\begin{bmatrix} X \\ XT \\ \vdots \\ XT^{j-1} \end{bmatrix}T\boldsymbol{x} = E\begin{bmatrix} XT \\ XT^2 \\ \vdots \\ XT^{j} \end{bmatrix}\boldsymbol{x} = \boldsymbol{0},$$

即 $\boldsymbol{x}\in\mathrm{Ker}(XT^{j+1})$. 按 $\mathscr{K}_{j+2}$的定义，$\boldsymbol{x}\in\mathscr{K}_{j+2}$，因而 $\mathscr{K}_{j+1}=\mathscr{K}_{j+2}$. 仿此继续下去便得 $\mathscr{K}_{j+1}=\mathscr{K}_{j+2}=\mathscr{K}_{j+3}=\cdots$. □

有了引理 1，我们可将满足 $\mathscr{K}_j=\mathscr{K}_{j+1}$ 的正整数 j 中最小者定义为 p 阶容许对 (X,T)的**指标**. 因此，当这个指标为 s 时，若记 $\dim\mathscr{K}_j=k_j$，$j=1,2,\cdots$，则由(3)式有 $k_1>k_2>\cdots>k_s=k_{s+1}=\cdots$. 因为这里的头 $s-1$ 个不等式关系严格成立且 $k_1\leqslant p$，所以一定有 $s\leqslant p$. 通常，将对应于上述指标为 s 的序列(3)中的子空间 $\mathscr{K}_s$ 称为**剩余子空间**(residual subspace). 显然，它包含在 $\mathrm{Ker}X$ 内，当 $\mathscr{K}_s=\{\boldsymbol{0}\}$，即 $k_s=0$ 时，p 阶容许对(X,T)称为**可观测的**(observable)，否则，剩余子空间 $\mathscr{K}_s$ 称为**非可观测的**，这时，(X,T)称为**非可观测的**. 我们将在后面解释这些概念在系统理论中的涵义.

下面定理 2 给出 p 阶容许对(X,T)为可观测的一个常用的充要条件. 注意，在此定理中，(X,T)的指标的值不直接出现.

定理 2　设(X,T)为 p 阶容许对. 则当且仅当

$$\mathrm{rank}\begin{bmatrix} X \\ XT \\ \vdots \\ XT^{p-1} \end{bmatrix} = p \tag{7}$$

时，(X,T)为可观测的.

证明　若条件(7)成立，按第一章 1.1.1 小节定理 6，$\dim\mathscr{K}_p=p-p=0$，即

$\mathscr{K}_p=\{\mathbf{0}\}$. 因此，按 p 阶容许对(X,T)指标 s 的定义，$\dim\mathscr{K}_p=\dim\mathscr{K}_s=0$（因为 $s\leqslant p$），因而(X,T)为可观测的. 反之，若(X,T)可观测，则由于其指标 $s\leqslant p$，故有 $\mathscr{K}_s=\mathscr{K}_p=\{\mathbf{0}\}$. 再次引用第一章 1.1.1 小节定理 6 便得条件(7). □

应用前面得到的结论与它们的对偶形式，可以导出可控的矩阵对的概念及其与可观测性之间的对偶关系.

对应于由(4)式定义的零空间 $\mathscr{K}_1,\mathscr{K}_2,\cdots$，令 $\mathscr{R}_l=\sum_{k=0}^{l-1}\operatorname{Im}((T^*)^kX^*)$，或等价地，

$$\mathscr{R}_l=\operatorname{Im}(X^*\quad T^*X^*\ \cdots\ T^{*l-1}X^*),\qquad l=1,2,\cdots. \tag{8}$$

显然，$\mathscr{R}_l$ 为$\mathbb{C}^p$ 的子空间，且

$$\mathscr{R}_1\subset\mathscr{R}_2\subset\mathscr{R}_3\subset\cdots.$$

应用第一章 1.2.1 小节(9′)式，我们有

$$\operatorname{Ker}\begin{bmatrix}X\\XT\\\vdots\\XT^{l-1}\end{bmatrix}\oplus\operatorname{Im}(X^*\ T^*X^*\ \cdots\ T^{*l-1}X^*)=\mathbb{C}^p,$$

亦即

$$\mathscr{K}_l\oplus\mathscr{R}_l=\mathbb{C}^p,\qquad l=1,2,\cdots. \tag{9}$$

因此，假如 s 为 p 阶容许对(X,T)的指标，则(9)式蕴涵

$$\mathscr{R}_1\subset\mathscr{R}_2\subset\cdots\subset\mathscr{R}_s=\mathscr{R}_{s+1}=\cdots,$$

式中"$\subset$"为严格包含关系，即有 $\mathscr{R}_1\subsetneqq\mathscr{R}_2$，等等. 并且，若 p 阶容许对(X,T)为可观测的($\dim\mathscr{K}_s=0$)，则由(9)式得到 $\mathscr{R}_s=\mathbb{C}^p$，即 $\sum_{k=0}^{s-1}\operatorname{Im}((T^*)^kX^*)=\mathbb{C}^p$.

通常，将可观测的矩阵对(X,T)的共轭矩阵对(T^*,X^*)称为**可控的**. 这时，$\mathscr{K}_s=\{\mathbf{0}\}$与 $\mathscr{R}_s=\mathbb{C}^p$. 假如 $\mathscr{R}_s\neq\mathbb{C}^p$ ($\mathscr{K}_s\neq\{\mathbf{0}\}$)，则称矩阵对$(T^*,X^*)$为**非可控的**. 它对应着矩阵对$(X,T)$的非可观测性.

应用(8)式与 $s\leqslant p$ 的事实，我们有

$$\mathscr{R}_s=\sum_{k=0}^{s-1}\operatorname{Im}((T^*)^kX^*)=\sum_{k=0}^{p-1}\operatorname{Im}((T^*)^kX^*), \tag{10}$$

它叫作矩阵对(T^*,X^*)的**可控的子空间**，记为 $\mathscr{C}_{T^*,X^*}$. 显然，它是(8)式诸 $\mathscr{R}_l$ 中$\mathbb{C}^p$ 内最大的 T^*-不变子空间.

综上所述，我们可以得到定理 2 的下列重要推论.

推论 3　设 $A\in\boldsymbol{M}_p(\mathbb{C})$与 $B\in\boldsymbol{M}_{p,r}(\mathbb{C})$. 则下列诸命题彼此等价：

(1)矩阵对(A,B)可控.

(2)矩阵对(B^*,A^*)可观测.

(3)$\operatorname{rank}(B\ AB\ \cdots\ A^{p-1}B)=p$.

(4)$\mathbb{C}^p=\sum_{k=0}^{p-1}\mathrm{Im}(A^kB)=\mathscr{C}_{A,B}$.

现在返回到受控的线性系统(1)～(2). 先给出此系统可观测的概念.

定义 4　系统(1)～(2)称为**可观测的**,假如对任意给定的控制 $\boldsymbol{u}(t)\in\mathscr{U}$,总存在有限的 $t_1>0$,使得系统的初始状态 $\boldsymbol{x}(0)=\boldsymbol{x}_0$ 能由区间$[0,t_1]$上的控制 $\boldsymbol{u}(t)$与观测 $\boldsymbol{y}(t)$的值唯一地确定.

因此,对于可观测的系统(1)～(2),当给定控制 $\boldsymbol{u}(t)\in\mathscr{U}$ 时,应用$[0,t_1]$上观测 $\boldsymbol{y}(t)$便可唯一地确定初始状态 $\boldsymbol{x}(0)=\boldsymbol{x}_0$,再通过求解初值问题:$\dot{\boldsymbol{x}}(t)=A\boldsymbol{x}(t)+B\boldsymbol{u}(t)$, $\boldsymbol{x}(0)=\boldsymbol{x}_0$,便求得 $\boldsymbol{x}(t)$在$[0,t_1]$上的向量值:

$$\boldsymbol{x}(t)=\mathrm{e}^{At}\boldsymbol{x}_0+\int_0^t\mathrm{e}^{A(t-\tau)}B\boldsymbol{u}(\tau)\mathrm{d}\tau,\qquad 0\leqslant t\leqslant t_1. \tag{11}$$

系统(1)～(2)的可观测性与矩阵对(C,A)的可观测性之间有简单的等价关系,确切地有

定理 5　设 $C\in\boldsymbol{M}_{r,n}(\mathbb{C})$与 $A\in\boldsymbol{M}_n(\mathbb{C})$组成 n 阶容许矩阵对(C,A). 则系统(1)～(2)可观测等价于矩阵对(C,A)可观测.

证明　设系统(1)～(2)可观测. 则对任意给定 $\boldsymbol{u}(t)\in\mathscr{U}$,有 $t_1>0$ 使得 $\boldsymbol{x}(0)=\boldsymbol{x}_0$(因而由(11)式确定的 $\boldsymbol{x}(t)(0\leqslant t\leqslant t_1)$)可由 $\boldsymbol{u}(t)$与 $\boldsymbol{y}(t)$在$[0,t_1]$上的值唯一确定. 这时,实际观测到的系统输出为

$$\boldsymbol{y}(t)=C\mathrm{e}^{At}\boldsymbol{x}_0+\int_0^tC\mathrm{e}^{A(t-\tau)}B\boldsymbol{u}(\tau)\mathrm{d}\tau,\qquad 0\leqslant t\leqslant t_1,$$

于是,矩阵方程

$$C\mathrm{e}^{At}\boldsymbol{x}_0=\boldsymbol{y}(t)-\int_0^tC\mathrm{e}^{A(t-\tau)}B\boldsymbol{u}(\tau)\mathrm{d}\tau,\qquad 0\leqslant t\leqslant t_1,$$

有唯一解 $\boldsymbol{x}_0$. 这等价于不存在 $\boldsymbol{x}_0\neq\boldsymbol{0}$ 使得 $C\mathrm{e}^{At}\boldsymbol{x}_0=\boldsymbol{0}$, $0\leqslant t\leqslant t_1$. 余下只要证明:这个最后条件相当于矩阵对$(C,A)$的可观测性. 事实上,若有 $\boldsymbol{x}_0\neq\boldsymbol{0}$ 使得 $C\mathrm{e}^{At}\boldsymbol{x}_0\equiv\boldsymbol{0}$, $0\leqslant t\leqslant t_1$,则对 $C\mathrm{e}^{At}\boldsymbol{x}_0$ 重复求 t 的导数便得 $CA\mathrm{e}^{At}\boldsymbol{x}_0\equiv\boldsymbol{0}$, $CA^2\mathrm{e}^{At}\boldsymbol{x}_0\equiv\boldsymbol{0}$, $\cdots$, $CA^{n-1}\mathrm{e}^{At}\boldsymbol{x}_0\equiv\boldsymbol{0}$, $0\leqslant t\leqslant t_1$,因而我们有 $Q(C,A)\mathrm{e}^{At}\boldsymbol{x}_0\equiv\boldsymbol{0}$, $0\leqslant t\leqslant t_1$,这里,

$$Q(C,A)=\begin{bmatrix}C\\CA\\\vdots\\CA^{n-1}\end{bmatrix}\in M_{nr,n}(\mathbb{C}).$$

取 $t=0$ 便知(C,A)的剩余子空间 $\mathrm{Ker}Q(C,A)\neq\{\boldsymbol{0}\}$,于是,矩阵对$(C,A)$非可观测. 反之,若矩阵对$(C,A)$非可观测,则有 $\mathrm{Ker}Q(C,A)\neq\{\boldsymbol{0}\}$,因而有 $\boldsymbol{x}_0\neq\boldsymbol{0}$ 使得 $CA^j\boldsymbol{x}_0=\boldsymbol{0}$, $0\leqslant j\leqslant n-1$. 按习题 6 的结果,我们有 $C\mathrm{e}^{At}\boldsymbol{x}_0=\boldsymbol{0}$, $0\leqslant t\leqslant t_1$.　□

从上述定理证明得出,若(C,A)为 n 阶容许矩阵对,且给定 $t_1>0$,则

$$\{\boldsymbol{x}\in\mathbb{C}^n: C\mathrm{e}^{At}\boldsymbol{x}=\boldsymbol{0},\ \forall t\in[0,t_1]\}=\bigcap_{j=0}^{n-1}\mathrm{Ker}(CA^j). \tag{12}$$

我们指出,定理 5 说明,系统(1)~(2)的可观测性与矩阵 B 的选取无关. 取 $B=O$,则系统(1)~(2)可观测当且仅当非零的初始状态 $\boldsymbol{x}_0$ 一定产生非零的输出 $\boldsymbol{y}(t)$. 在这种意义下,我们认为系统有可观测到的解.

现在考察受控线性系统(1)的可控性问题(从后面推论 10 看出,系统(1)~(2)的可控性不必涉及到方程(2)). 给定 $\boldsymbol{x}(0)=\boldsymbol{x}_0$ 与控制 $\boldsymbol{u}(t)\in\mathscr{U}$,初值问题 $\dot{\boldsymbol{x}}(t)=A\boldsymbol{x}(t)+B\boldsymbol{u}(t)$, $\boldsymbol{x}(0)=\boldsymbol{x}_0$ 有唯一解

$$\boldsymbol{x}(t;\boldsymbol{x}_0,\boldsymbol{u}) = \mathrm{e}^{At}\boldsymbol{x}_0 + \int_0^t \mathrm{e}^{A(t-\tau)}B\boldsymbol{u}(\tau)\mathrm{d}\tau. \tag{11'}$$

定义 6 向量 $\boldsymbol{y}\in\mathbb{C}^n$ 叫作系统(1)从初始向量 $\boldsymbol{x}_0\in\mathbb{C}^n$ **可达的状态**,假如存在 $t>0$ 与 $\boldsymbol{u}\in\mathscr{U}$ 使得 $\boldsymbol{x}(t;\boldsymbol{x}_0,\boldsymbol{u})=\boldsymbol{y}$. 这时我们说,$\boldsymbol{y}$ 为从 $\boldsymbol{x}_0$ **经过时间 t 后可达的状态**. 假如 $\mathbb{C}^n$ 中每一个向量都是系统(1)从 $\mathbb{C}^n$ 中另外向量可达的状态,则称系统(1)为**可控的**.

在后面推论 10 中,将证明系统(1)的可控性等价于矩阵对 (A,B) 的可控性. 为此先考虑一下向量 $\boldsymbol{y}\in\mathbb{C}^n$ 为系统(1)从 $\boldsymbol{z}$(经过时间 t 后)可达的状态的条件.

定理 7 设 $\boldsymbol{y},\boldsymbol{z}\in\mathbb{C}^n$. 则当且仅当

$$\boldsymbol{y} - \mathrm{e}^{At}\boldsymbol{z} \in \mathscr{C}_{A,B} \tag{13}$$

时,向量 $\boldsymbol{y}$ 是系统(1)从 $\boldsymbol{z}$(经过时间 t 后)可达的状态. 这里,$\mathscr{C}_{A,B}=\sum_{r=0}^{n-1}\operatorname{Im}(A^rB)$ 表示矩阵对 (A,B) 的可控的子空间.

我们先研究一个引理,它给出矩阵对 (A,B) 的可控的子空间 $\mathscr{C}_{A,B}$ 的一个有用的表达式. 待证明此引理的一个推论之后再返回来证明定理 7.

引理 8 设矩阵值函数 $W(t)$ 为

$$W(t) = \int_0^t \mathrm{e}^{A\tau}BB^*\mathrm{e}^{A^*\tau}\mathrm{d}\tau. \tag{14}$$

则对任意 $t>0$, $\operatorname{Im}W(t)=\mathscr{C}_{A,B}$.

证明 显然,对 $t>0$, $W(t)\in\boldsymbol{M}_n(\mathbb{C})$ 有定义,且 $W(t)=W(t)^*$,因而我们有 $\operatorname{Im}W(t)=(\operatorname{Ker}W(t))^{\perp}$(见第一章习题 1.2.1 题 6). 用 $\mathscr{K}_{A,B}$ 表示矩阵对 (B^*,A^*) 的非可观测的子空间,则有 $\mathscr{C}_{A,B}\oplus\mathscr{K}_{A,B}=\mathbb{C}^n$,因而 $\mathscr{C}_{A,B}=\mathscr{K}_{A,B}^{\perp}$. 这样一来,本引理的结论相当于: $\operatorname{Ker}W(t)=\mathscr{K}_{A,B}$, $\forall t>0$. 现在来证明这个等式. 若 $\boldsymbol{x}\in\operatorname{Ker}W(t)$, $t>0$,则

$$\boldsymbol{x}^*W(t)\boldsymbol{x} = \int_0^t \boldsymbol{x}^*\mathrm{e}^{A\tau}BB^*\mathrm{e}^{A^*\tau}\boldsymbol{x}\mathrm{d}\tau = \int_0^t \|B^*\mathrm{e}^{A^*\tau}\boldsymbol{x}\|_2^2\mathrm{d}\tau = 0,$$

式中,$\|\cdot\|_2$ 是 $\mathbb{C}^n$ 上欧氏向量范数. 因此,$B^*\mathrm{e}^{A^*\tau}\boldsymbol{x}=\boldsymbol{0}$, $\forall\tau\in[0,t]$. 反之,若 $\boldsymbol{x}\in\mathbb{C}^n$ 满足 $B^*\mathrm{e}^{A^*\tau}\boldsymbol{x}=\boldsymbol{0}$, $\forall\tau\in[0,t]$,则由(14)式看出,$\boldsymbol{x}\in\operatorname{Ker}W(t)$, $\forall t>0$. 因此,按(12)式,$\boldsymbol{x}\in\operatorname{Ker}W(t)$, $\forall t>0$ 的事实等价于: $\boldsymbol{x}\in\bigcap_{j=1}^{n}\operatorname{Ker}(B^*(A^*)^{j-1})=\mathscr{K}_{A,B}$. 于是,$\mathscr{K}_{AB}=\operatorname{Ker}W(t)$, $\forall t>0$. □

从引理 8 可导出矩阵对 (A,B) 可控性的另一个重要的等价原则.

推论 9　当且仅当(14)式中矩阵 $W(t)$ 对所有 $t>0$ 为 Hermite 正定时,矩阵对 (A,B) 为可控的.

证明　由(14)式看出,对所有 $t>0$,$W(t)$ 为 Hermite 正半定的. 按推论 3,矩阵对 (A,B) 可控性等价于 $\mathscr{C}_{A,B}=\mathbb{C}^n$. 但这时引理 8 蕴涵 $\mathbb{C}^n=\mathscr{C}_{A,B}=\mathrm{Im}W(t)$, $\forall t>0$,此表明对所有 $t>0$, $\mathrm{rank}W(t)=n$. 因此,矩阵对 (A,B) 的可控性相当于:对所有 $t>0$, $W(t)$ 为 Hermite 正定的. □

现在返回到定理 7 的证明.

定理 7 的证明　设 $\boldsymbol{y}\in\mathbb{C}^n$ 为系统(1)从 $\boldsymbol{z}\in\mathbb{C}^n$(经过时间 t 后)可达的状态. 按(11′)式有

$$\boldsymbol{y}-\mathrm{e}^{At}\boldsymbol{z}=\int_0^t \mathrm{e}^{A(t-\tau)}B\boldsymbol{u}(\tau)\mathrm{d}\tau,$$

式中,$\boldsymbol{u}$ 为 $\mathscr{U}$ 中某个控制. 应用习题 6 的结论,上式可改写为

$$\boldsymbol{y}-\mathrm{e}^{At}\boldsymbol{z}=\sum_{j=1}^n A^{j-1}B\int_0^t \psi_j(t-\tau)\boldsymbol{u}(\tau)\mathrm{d}\tau,$$

这里诸 ψ_j 为整函数. 由于矩阵对 (A,B) 的可控的子空间为 $\mathscr{C}_{A,B}=\sum\limits_{j=1}^n \mathrm{Im}(A^{j-1}B)$,故上述最后式子意味着 $\boldsymbol{y}-\mathrm{e}^{At}\boldsymbol{z}\in\mathscr{C}_{A,B}$. 反之,设(13)式成立. 依照可达的状态的定义,我们只要证明存在 $\boldsymbol{u}\in\mathscr{U}$ 使得 $\boldsymbol{y}=\boldsymbol{x}(t;\boldsymbol{z},\boldsymbol{u})$. 按(11′)式这相当于

$$\boldsymbol{y}-\mathrm{e}^{At}\boldsymbol{z}=\int_0^t \mathrm{e}^{A(t-\tau)}B\boldsymbol{u}(\tau)\mathrm{d}\tau.$$

但应用引理 8,存在 $\boldsymbol{w}\in\mathbb{C}^n$ 使得

$$\boldsymbol{y}-\mathrm{e}^{At}\boldsymbol{z}=W(t)\boldsymbol{w}=\int_0^t \mathrm{e}^{A\tau}BB^*\mathrm{e}^{A^*\tau}\boldsymbol{w}\mathrm{d}\tau.$$

现令 $\tau=t-\sigma$,由上式推出

$$\boldsymbol{y}-\mathrm{e}^{At}\boldsymbol{z}=\int_0^t \mathrm{e}^{A(t-\sigma)}BB^*\mathrm{e}^{A^*(t-\sigma)}\boldsymbol{w}\mathrm{d}\sigma.$$

于是,取 $\boldsymbol{u}(\tau)=B^*\mathrm{e}^{A^*(t-\tau)}\boldsymbol{w}$ 作为前面的控制即可,显然,此 $\boldsymbol{u}\in\mathscr{U}$. □

作为定理 7 的推论,我们有下列重要的结论.

推论 10　系统(1)是可控的充分与必要条件为矩阵对 (A,B) 是可控的.

证明　若矩阵对 (A,B) 可控,则按推论 3,$\mathscr{C}_{A,B}=\sum\limits_{r=0}^{n-1}\mathrm{Im}(A^rB)=\mathbb{C}^n$. 因此,对任意的 $\boldsymbol{y},\boldsymbol{z}\in\mathbb{C}^n$,显然有 $\boldsymbol{y}-\mathrm{e}^{At}\boldsymbol{z}\in\mathscr{C}_{A,B}$. 按定理 7,$\boldsymbol{y}$ 是系统(1) 从 $\boldsymbol{z}$ 可达的状态. 由于 $\boldsymbol{y}$ 与 $\boldsymbol{z}$ 的随意性,我们推得系统(1) 的可控性. 反之,若系统(1) 可控,这时特别有:每个 $\boldsymbol{y}\in\mathbb{C}^n$ 是系统(1) 从零向量 $\boldsymbol{0}\in\mathbb{C}^n$ 可达的状态. 按定理 7,这时 $\boldsymbol{y}\in\mathscr{C}_{A,B}$,于是,$\mathbb{C}^n=\mathscr{C}_{A,B}$. 根据推论 3,这表明矩阵对 (A,B) 为可控的. □

我们强调指出,若系统(1)可控,那么,任意 $\boldsymbol{y}\in\mathbb{C}^n$ 可视为系统(1)从任意 $\boldsymbol{z}\in$

$\mathbb{C}^n$ 经过**任意时间** $t>0$ 后可达的状态. 事实上,对任意给定的 $t>0$,按推论 9 可取如下的控制 $\boldsymbol{u}\in\mathscr{U}$:

$$\boldsymbol{u}(\tau)=B^*\mathrm{e}^{A^*(t-\tau)}\boldsymbol{W}(t)^{-1}(\boldsymbol{y}-\mathrm{e}^{At}\boldsymbol{z}). \tag{15}$$

这时,(11′)与(14)式给出

$$\begin{aligned}\boldsymbol{x}(t;\boldsymbol{z},\boldsymbol{u})&=\mathrm{e}^{At}\boldsymbol{z}+\left(\int_0^t\mathrm{e}^{A(t-\tau)}BB^*\mathrm{e}^{A^*(t-\tau)}\mathrm{d}\tau\right)\boldsymbol{W}(t)^{-1}(y-\mathrm{e}^{At}\boldsymbol{z})\\&=\mathrm{e}^{At}\boldsymbol{z}+\boldsymbol{y}-\mathrm{e}^{At}\boldsymbol{z}=\boldsymbol{y}.\end{aligned}$$

在系统理论里,通常将如下线性系统:

$$\begin{aligned}\dot{\boldsymbol{y}}(t)&=A^*\boldsymbol{y}(t)\\\boldsymbol{z}(t)&=B^*\boldsymbol{y}(t)\end{aligned} \tag{16}$$

称为系统(1)的**对偶系统**. 按推论 10,系统(1)的可控性相当于矩阵对(A,B)的可控性,但后者又相当于矩阵对(B^*,A^*)的可观测性. 再引用定理 5,这些条件也等价于系统(16)的可观测性. 因此,我们有

定理 11　系统(1)的可控性等价于它的对偶系统(16)的可观测性.

这个定理通常叫作**对偶原理**. 它从理论上阐明了系统的可控性问题与可观测性问题是彼此对偶的,即任何系统(1)的可控性问题总可归结为另一与之对偶的系统的可观测性问题,反之亦然.

最后,我们讨论一个质点运动的受控线性系统的例子作为本小节的总结.

例 12　考虑质量为 m 的质点沿水平轨道的直线运动. 若用 $x(t)$ 记质点在时刻 t 的位置,$u(t)$ 记施于质点的外部水平控制力,它逐段连续,则该质点的运动方程(不计空气阻力)为

$$\begin{aligned}m\ddot{x}(t)&=u(t),\qquad t>0\\x(0)&=x_{00},\qquad \dot{x}(0)=x_{10},\end{aligned}$$

其中,常数 x_{00} 与 x_{10} 分别为质点的初始位置与初速. 现令 $x_1(t)=x(t)$, $x_2(t)=\dot{x}_1(t)$, $\boldsymbol{x}(t)=(x_1(t),x_2(t))^{\mathrm{T}}$, $A=\begin{bmatrix}0&1\\0&0\end{bmatrix}$与 $B=\begin{bmatrix}0\\m^{-1}\end{bmatrix}$, $\boldsymbol{x}_0=(x_{00},x_{10})^{\mathrm{T}}$,则上述运动方程可改写为

$$\begin{aligned}\dot{\boldsymbol{x}}(t)&=A\boldsymbol{x}(t)+Bu(t),\\\boldsymbol{x}(0)&=\boldsymbol{x}_0.\end{aligned} \tag{17}$$

从直观容易看出,系统(17)是可控的,因为任意 $\boldsymbol{y}=(y_1,y_2)^{\mathrm{T}}\in\mathbb{R}^2$ 总是系统(17)从任一其他向量 $\boldsymbol{z}=(z_1,z_2)^{\mathrm{T}}\in\mathbb{R}^2$ 的可达的状态,也就是说,质点 m 位于 y_1 处具有速度 y_2 的状态总可以从质点 m 位于任意 z_1 处具有任意速度 z_2 的状态在适当外力 $u(t)$ 驱使下达到. 现在应用前面的知识,用不同方法从理论上判定系统(17)的可控性.

应用推论 3 与 10,经过简单计算可得

$$\operatorname{rank}[B \quad AB] = \operatorname{rank}\begin{bmatrix} 0 & m^{-1} \\ m^{-1} & 0 \end{bmatrix} = 2,$$

因而矩阵对(A,B)是可控的,此等价于系统(17)的可控性.

假如应用推论 9,经过简单计算可得

$$e^{A\tau}B = \begin{bmatrix} 1 & \tau \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 0 \\ m^{-1} \end{bmatrix} = \begin{bmatrix} \tau m^{-1} \\ m^{-1} \end{bmatrix},$$

$$\begin{aligned} W(t) &= \int_0^t e^{A\tau}BB^* e^{A^*\tau} d\tau = \int_0^t \begin{bmatrix} \tau m^{-1} \\ m^{-1} \end{bmatrix} [\tau m^{-1} m^{-1}] d\tau \\ &= \int_0^t m^{-2} \begin{bmatrix} \tau^2 & \tau \\ \tau & 1 \end{bmatrix} d\tau \\ &= m^{-2} \begin{bmatrix} \frac{t^3}{3} & \frac{t^2}{2} \\ \frac{t^2}{2} & t \end{bmatrix}. \end{aligned}$$

显然,对任意$t>0$,$W(t)$为 Hermite 正定的. 因此,矩阵对(A,B)是可控的——再次证实系统(17)的可控性.

根据(15)式,假如选取控制$u(\tau)$为

$$\begin{aligned} u(\tau) &= B^* e^{A^*(t-\tau)} W(t)^{-1} (\boldsymbol{y} - e^{At}\boldsymbol{z}) \\ &= \frac{12m}{t^3}\left[\left(z_1 - y_1 + \frac{t}{2}y_2 + \frac{t}{2}z_2\right)\tau + \frac{t}{2}(y_1 - z_1) - \frac{t^2}{6}y_2 - \frac{t^2}{3}z_2\right], \end{aligned}$$

则质点m从状态$\boldsymbol{z}=(z_1,z_2)^{\mathrm{T}}$经过时间$t$后到达状态$\boldsymbol{y}=(y_1,y_2)^{\mathrm{T}}$.

现在考虑系统(17)的对偶系统:

$$\begin{bmatrix} \dot{y}_1(t) \\ \dot{y}_2(t) \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix},$$

$$z(t) = [0 \quad m^{-1}]\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = m^{-1}y_2(t).$$

它是可观测的. 除了应用定理 11 或定理 5 以外,也可以应用系统可观测性定义直接验证这个事实. 实际上,由于$z(t)=m^{-1}y_2(t)$,故$y_2(0)$完全由$z(0)$唯一确定. 同时,由于$y_2(t)=y_1(0)t+y_2(0)=mz(t)$,故有$y_1(0)=\frac{1}{t}(mz(t)-y_2(0))$,因而$y_1(0)$也由$[0,t](t>0)$上的$z(t)$值唯一确定. □

习题 3.4.2

1. 设(X,T)为p阶容许对. 试证:

(1)若$r=1$,$X\neq O$,且$T=I_p$,则(X,I_p)的指标为 1,且$k_1=k_2=\cdots=p-1$.

(2)若$r=1$,$X\neq O$,且(X,T)可观测,则(X,T)的指标为p.

(3)若 x^{T} 为 T 的左特征向量,则矩阵对 (x^{T},T) 有指标 1.

2. 试证:容许对 (X,T) 的非可观测子空间为 T-不变的.

3. 设 (X,T) 为 p 阶容许对, $X\in \boldsymbol{M}_{r,p}(\mathbb{C})$. 试证:对任意 $G\in \boldsymbol{M}_{p,r}(\mathbb{C})$,如下矩阵对

$$(X,T) \quad 与 \quad (X,T+GX)$$

有相同的指标与非可观测的子空间.

4. 设 $A\in \boldsymbol{M}_n(\mathbb{C})$ 与 $B\in \boldsymbol{M}_{n,m}(\mathbb{C})$. 试证下列诸命题等价:

(1)矩阵对 (A,B) 可控.

(2)对任意 $K\in \boldsymbol{M}_{m,n}(\mathbb{C})$,矩阵对 $(A+BK,B)$ 可控.

(3)对任意复数 λ,矩阵对 $(A-\lambda I,B)$ 可控.

5. 设 $\mathscr{C}_{A,B}$ 维数为 k. 试求矩阵 P 使得如果 $\hat{A}=P^{-1}AB$ 与 $\hat{B}=P^{-1}B$,则

$$\hat{A}=\begin{bmatrix} A_1 & A_3 \\ O & A_2 \end{bmatrix} \quad 与 \quad \hat{B}=\begin{bmatrix} B_1 \\ O \end{bmatrix},$$

其中, $A_1\in \boldsymbol{M}_k(\mathbb{C})$, $B_1\in \boldsymbol{M}_{k,m}(\mathbb{C})$,且矩阵对 (A_1,B_1) 可控.

6. 试证:对任意 $A\in \boldsymbol{M}_n(\mathbb{C})$,存在整函数 $\psi_1(t),\cdots,\psi_n(t)$ 使得

$$\mathrm{e}^{At}=\sum_{j=1}^{n}\psi_j(t)A^{j-1}.$$

7. 试证:若 $A\in \boldsymbol{M}_n(\mathbb{C})$,则存在复值函数 $\varphi_1(\lambda),\cdots,\varphi_n(\lambda)$,它们在 A 的预解集内解析,使得

$$(\lambda I-A)^{-1}=\sum_{j=1}^{n}\varphi_j(\lambda)A^{j-1}.$$

8. 设 (C,A) 为 n 阶容许矩阵对. 试证: (C,A) 可观测,当且仅当不存在 $x_0\neq \mathbf{0}$ 使得对充分大的 $|\lambda|$,都有 $C(\lambda I-A)^{-1}x_0=\mathbf{0}$.

9. 试证:矩阵对 (A,B) 的可控的子空间 $\mathscr{C}_{A,B}$ 为包含 $\mathrm{Im}B$ 的最小 A-不变子空间.

参考文献

[1] Lancaster P, Tismenetsky M. The Theory of Matrices, with Applications. 2nd ed. New York: Academic Press, 1985

[2] Rudin W. Functional Analysis. New York: McGraw-Hill, Inc. , 1973

[3] Hoffman K, Kunze R. Linear Algebra. 2nd ed. Englewood Cliffs: Prentice-Hall, N. J. 1971

[4] Hirsch M W, Smale S. Differential Equations, Dynamical Systems, and Linear Algebra. Academic. New York: Preess, 1974

[5] Sontag E D. Mathematical Control Theory 2nd ed. New York: Springer, 1998

[6] Dunford N, Schwartz J T. Linear Operators. New York: Pt. I, Interscience, 1958

[7] Barnett S. Polynomials and linear Control systems. New York: Marcel Dekker, 1983

[8] Graham A. Kronecker Products and Matrix Calculus with Applications. Chichester, U. K. Horwood, 1981

[9] Russell D L. Mathematics of Finite-Dimensional Control Systems. New York. Marcel Dekker, 1979

第四章　线性矩阵方程与惯性理论

本章讨论复矩阵的特征值相对于虚轴、实轴或单位圆周的分布，即**矩阵惯性理论**，与多项式的零点相对于虚轴、实轴或单位圆周的分布，即**多项式惯性理论**，以及与惯性理论有密切联系的线性矩阵方程. 两种特殊的线性矩阵方程：Ляпунов 方程与 Stein 方程，在稳定性问题中起着基本的作用. 同时，线性矩阵方程与矩阵广义逆理论之间也有着密切的联系(见第五章).

4.1　线性矩阵方程

4.1.1　矩阵的张量积

对于两个矩阵 $A=(a_{ij})\in \boldsymbol{M}_{m,n}(\boldsymbol{F})$ 与 $B=(b_{ij})\in \boldsymbol{M}_{p,q}(\boldsymbol{F})$，我们已经引入乘积 AB(这时要求 $n=p$)与 Hadamard 积 $A\circ B$(这时要求 $m=p$ 与 $n=q$). 现在介绍另一种所谓**张量积或 Kronecker 乘积**，记为 $A\otimes B$，它由如下 $mp\times nq$ 矩阵确定：

$$A\otimes B=[a_{ij}B]_{i,j=1}^{m,n}=\begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{bmatrix}. \tag{1}$$

因此，它可视为有 m 个块行与 n 个块列的分块矩阵，其各块均为一个 $p\times q$ 矩阵. 由此定义看出一般地 $A\otimes B\neq B\otimes A$，虽然二者有相同的大小.

例 1　(1)设 $A=(a_{ij})\in \boldsymbol{M}_{2,3}(\boldsymbol{F})$ 与 $B=(b_{ij})\in \boldsymbol{M}_2(\boldsymbol{F})$. 则有

$$A\otimes B=[a_{ij}B]_{i,j=1}^{2,3}=\begin{bmatrix} a_{11}b_{11} & a_{11}b_{12} & a_{12}b_{11} & a_{12}b_{12} & a_{13}b_{11} & a_{13}b_{12} \\ a_{11}b_{21} & a_{11}b_{22} & a_{12}b_{21} & a_{12}b_{22} & a_{13}b_{21} & a_{13}b_{22} \\ a_{21}b_{11} & a_{21}b_{12} & a_{22}b_{11} & a_{22}b_{12} & a_{23}b_{11} & a_{23}b_{12} \\ a_{21}b_{21} & a_{21}b_{22} & a_{22}b_{21} & a_{22}b_{22} & a_{23}b_{21} & a_{23}b_{22} \end{bmatrix},$$

$$B\otimes A=[b_{ij}A]_{i,j=1}^{2,2}=\begin{bmatrix} b_{11}a_{11} & b_{11}a_{12} & b_{11}a_{13} & b_{12}a_{11} & b_{12}a_{12} & b_{12}a_{13} \\ b_{11}a_{21} & b_{11}a_{22} & b_{11}a_{23} & b_{12}a_{21} & b_{12}a_{22} & b_{12}a_{23} \\ b_{21}a_{11} & b_{21}a_{12} & b_{21}a_{13} & b_{22}a_{11} & b_{22}a_{12} & b_{22}a_{13} \\ b_{21}a_{21} & b_{21}a_{22} & b_{21}a_{23} & b_{22}a_{21} & b_{22}a_{22} & b_{22}a_{23} \end{bmatrix}.$$

(2)应用有限差分方法求解长 $6h$ 宽 $4h$ 的平面矩形区域 Ω 的 Dirichlet 问题

$$\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}=0,\qquad (x,y)\in \Omega,$$

$$u|_{\Gamma} = g(x,y),$$

其中,Γ 为 Ω 的边界,g 为定义在 Γ 上的给定的函数. 在 Ω 上应用边长为 h 的正方形网格,其内节点排列如图 4.1 所示,则对应的一种差分方程组的系数矩阵 $M \in \boldsymbol{M}_{15}(\boldsymbol{F})$ 有如下形式

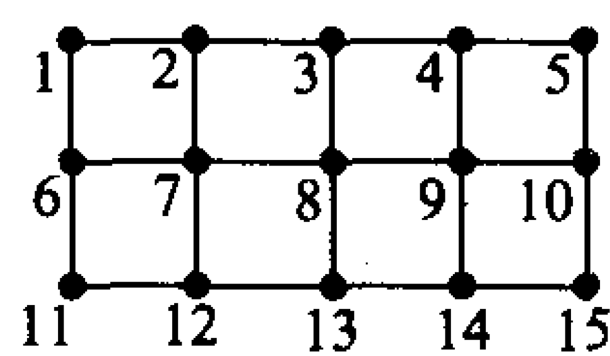


图 4.1

$$M = \frac{1}{4}\left[\begin{array}{ccccc|ccccc|ccccc}
4 & -1 & & & & -1 & & & & & & & & & \\
-1 & 4 & -1 & & & & -1 & & & & & & & & \\
 & -1 & 4 & -1 & & & & -1 & & & & & & & \\
 & & -1 & 4 & -1 & & & & -1 & & & & & & \\
 & & & -1 & 4 & & & & & -1 & & & & & \\
\hline
-1 & & & & & 4 & -1 & & & & -1 & & & & \\
 & -1 & & & & -1 & 4 & -1 & & & & -1 & & & \\
 & & -1 & & & & -1 & 4 & -1 & & & & -1 & & \\
 & & & -1 & & & & -1 & 4 & -1 & & & & -1 & \\
 & & & & -1 & & & & -1 & 4 & & & & & -1 \\
\hline
 & & & & & -1 & & & & & 4 & -1 & & & \\
 & & & & & & -1 & & & & -1 & 4 & -1 & & \\
 & & & & & & & -1 & & & & -1 & 4 & -1 & \\
 & & & & & & & & -1 & & & & -1 & 4 & -1 \\
 & & & & & & & & & -1 & & & & -1 & 4
\end{array}\right]. \tag{2}$$

容易验证,矩阵 M 可表示为

$$M = (I_3 \otimes I_5) - \frac{1}{4}(Z_3 \otimes I_5) - \frac{1}{4}(I_3 \otimes Z_5), \tag{3}$$

其中,

$$Z_3 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad Z_5 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

下面考虑矩阵张量积的基本性质.

定理 2 设矩阵 A,B,C,D 使得 AC 与 BD 有定义. 则

$$(A \otimes B)(C \otimes D) = (AC) \otimes (BD). \tag{4}$$

证明 设 A 为 $m \times n$ 矩阵,C 为 $n \times r$ 矩阵. 则(4)式左端等于

$$\begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{bmatrix} \begin{bmatrix} c_{11}D & c_{12}D & \cdots & c_{1r}D \\ c_{21}D & c_{22}D & \cdots & c_{2r}D \\ \vdots & \vdots & & \vdots \\ c_{n1}D & c_{n2}D & \cdots & c_{nr}D \end{bmatrix}.$$

按分块矩阵乘法运算,上式等于

$$\begin{bmatrix} (a_{11}c_{11}+a_{12}c_{21}+\cdots+a_{1n}c_{n1})BD & \cdots & (a_{11}c_{1r}+a_{12}c_{2r}+\cdots+a_{1n}c_{nr})BD \\ \vdots & & \vdots \\ (a_{m1}c_{11}+a_{m2}c_{21}+\cdots+a_{mn}c_{n1})BD & \cdots & (a_{m1}c_{1r}+a_{m2}c_{2r}+\cdots+a_{mn}c_{nr})BD \end{bmatrix}$$

$$=(AC)\otimes(BD).$$ □

从定义与上述定理我们有一些有用的推论.

推论 3　设 $A\in \boldsymbol{M}_m(\boldsymbol{F})$ 与 $B\in \boldsymbol{M}_n(\boldsymbol{F})$ 皆为非奇异矩阵. 则 $A\otimes B\in \boldsymbol{M}_{mn}(\boldsymbol{F})$ 也是非奇异矩阵,并且,

$$(A\otimes B)^{-1}=A^{-1}\otimes B^{-1}. \tag{5}$$

证明　按(4)式我们有

$$(A\otimes B)(A^{-1}\otimes B^{-1})=(AA^{-1})\otimes(BB^{-1})=I_m\otimes I_n=I_{mn}.$$ □

推论 4　若 A 与 B 为复矩阵,则

$$(A\otimes B)^{\mathrm{T}}=A^{\mathrm{T}}\otimes B^{\mathrm{T}},\overline{(A\otimes B)}=\overline{A}\otimes\overline{B},\qquad (A\otimes B)^*=A^*\otimes B^*.$$

推论 5　若 $A\in \boldsymbol{M}_m(\boldsymbol{F})$ 相似于 A_1,且有 $A=PA_1P^{-1}$,$B\in \boldsymbol{M}_n(\boldsymbol{F})$ 相似于 B_1,且有 $B=QB_1Q^{-1}$,则 $A\otimes B\in \boldsymbol{M}_{mn}(\boldsymbol{F})$ 相似于 $A_1\otimes B_1$,且有

$$A\otimes B=(P\otimes Q)(A_1\otimes B_1)(P\otimes Q)^{-1}. \tag{6}$$

推论 6　若 $A=(a_{ij})\in \boldsymbol{M}_m(\boldsymbol{F})$ 与 $B=(b_{ij})\in \boldsymbol{M}_n(\boldsymbol{F})$ 皆为上三角形矩阵,则 $A\otimes B\in \boldsymbol{M}_{mn}(\boldsymbol{F})$ 也是上三角形矩阵,且它的主对角元素依序为 $a_{11}b_{11},a_{11}b_{22},\cdots,a_{11}b_{nn}$, $a_{22}b_{11},a_{22}b_{22},\cdots,a_{22}b_{nn},\cdots,a_{mm}b_{11},a_{mm}b_{22},\cdots,a_{mm}b_{nn}$.

为了证明关于张量积特征值与特征向量的基本结果(见推论 8(1)),我们先考虑它的更一般的结果:**Stephanos 定理**. 设 $A\in \boldsymbol{M}_m(\mathbb{C})$,$B\in \boldsymbol{M}_n(\mathbb{C})$,且 p 为两个独立变量 x 与 y 的复系数多项式

$$p(x,y)=\sum_{i,j=0}^{r}c_{ij}x^iy^j,$$

定义

$$p(A;B)=\sum_{i,j=0}^{r}c_{ij}A^i\otimes B^j. \tag{7}$$

定理 7(Stephanos)　设 $A\in \boldsymbol{M}_m(\mathbb{C})$ 有特征值 $\alpha_1,\alpha_2,\cdots,\alpha_m$;$B\in \boldsymbol{M}_n(\mathbb{C})$ 有特征值 $\beta_1,\beta_2,\cdots,\beta_n$. 则 $p(A;B)\in \boldsymbol{M}_{mn}(\mathbb{C})$ 有特征值 $p(\alpha_k,\beta_l)$, $k=1,2,\cdots,m$; $l=1,2,\cdots,n$. 如果 $\boldsymbol{u}\in\mathbb{C}^m$ 为对应于 α_k 的 A 的特征向量,$\boldsymbol{v}\in\mathbb{C}^n$ 为对应于 β_l 的 B 的特征向量,则 $\boldsymbol{u}\otimes\boldsymbol{v}\in\mathbb{C}^{mn}$ 为对应于特征值 $p(\alpha_k,\beta_l)$ 的 $p(A;B)$ 的特征向量.

证明　设 $A=PJ_AP^{-1}$ 与 $B=QJ_BQ^{-1}$,其中,J_A 与 J_B 分别为 A 与 B 的 Jordan 正规形式. 则 J_A^i 是以 $\alpha_1^i,\cdots,\alpha_m^i$ 为对角元素的 m 阶上三角形矩阵与 J_B^i 是以 $\beta_1^i,\cdots,\beta_n^i$ 为对角元素的 n 阶上三角形矩阵,因而按张量积性质,$J_A^i\otimes J_B^i$ 也是上三角形矩阵,其对角元素为 $\alpha_k^i\beta_l^i$, $k=1,2,\cdots,m$; $l=1,2,\cdots,n$. 于是,$p(J_A;J_B)$

$=\sum_{i,j=0}^{r} c_{ij}(J_A^i \otimes J_B^j)$是 mn 阶上三角形矩阵，其对角元素为 $p(\alpha_k,\beta_l)$，$k=1,2,\cdots,m$；$l=1,2,\cdots,n$. 由推论 5 知道，$p(A;B)$相似于 $p(J_A;J_B)$，因而它们有相同的特征值.

现设 $A\boldsymbol{u}=\alpha_k\boldsymbol{u}(\boldsymbol{u}\neq\boldsymbol{0})$与 $B\boldsymbol{v}=\beta_l\boldsymbol{v}(\boldsymbol{v}\neq\boldsymbol{0})$，则(4)式保证

$$(A^i\otimes B^j)(\boldsymbol{u}\otimes\boldsymbol{v})=A^i\boldsymbol{u}\otimes B^j\boldsymbol{v}=\alpha_k^i\beta_l^j(\boldsymbol{u}\otimes\boldsymbol{v}),\quad i,j=0,1,\cdots,r.$$

同时，$\boldsymbol{u}\neq\boldsymbol{0}$ 与 $\boldsymbol{v}\neq\boldsymbol{0}$ 蕴涵 $\boldsymbol{u}\otimes\boldsymbol{v}\neq\boldsymbol{0}$. 由此推出，$\boldsymbol{u}\otimes\boldsymbol{v}$ 是对应于特征值 $p(\alpha_k,\beta_l)$的 $p(A;B)$的特征向量. □

取 $p(x,y)=xy$ 与 $p(x,y)=x+y$，定理 7 直接推出下列结论.

推论 8　设 $A\in\boldsymbol{M}_m(\mathbb{C})$与 $B\in\boldsymbol{M}_n(\mathbb{C})$有特征值与特征向量同定理 7. 则：

(1)$A\otimes B$ 与 $B\otimes A$ 的 mn 个特征值均为 $\alpha_k\beta_l$，$k=1,2,\cdots,m$；$l=1,2,\cdots,n$，且对应于 $\alpha_k\beta_l$ 的特征向量分别为 $\boldsymbol{u}\otimes\boldsymbol{v}$ 与 $\boldsymbol{v}\otimes\boldsymbol{u}$.

(2)$(A\otimes I_n)+(I_m\otimes B)$与$(I_n\otimes A)+(B\otimes I_m)$的 mn 个特征值均为 $\alpha_k+\beta_l$，$k=1,2,\cdots,m$；$l=1,2,\cdots,n$，且对应于 $\alpha_k+\beta_l$ 的特征向量分别为 $\boldsymbol{u}\otimes\boldsymbol{v}$ 与 $\boldsymbol{v}\otimes\boldsymbol{u}$.

通常，将$(I_n\otimes A)+(B\otimes I_m)\in\boldsymbol{M}_{mn}(\mathbb{C})$叫作 A 与 B 的**张量和**(或 **Kronecker 和**)，记为 $A\oplus B$. 按此记号，(3)式中矩阵 M 可以改写为 $I_{15}-\frac{1}{4}(Z_5\oplus Z_3)$，并且，假如我们已知 Z_3 的特征值为 $\alpha_1,\alpha_2,\alpha_3$；$Z_5$ 的特征值为 $\beta_1,\cdots,\beta_5$，则应用推论 8(2)的结果可得$M\in\boldsymbol{M}_{15}(\mathbb{C})$的特征值为 $1-\frac{1}{4}(\alpha_i+\beta_j)$，$1\leqslant i\leqslant 3$，$1\leqslant j\leqslant 5$.

应用定理 7 或推论 8 可得张量积的行列式，迹与秩的公式.

推论 9　设 $A\in\boldsymbol{M}_m(\mathbb{C})$与 $B\in\boldsymbol{M}_n(\mathbb{C})$. 则有

$$\det(A\otimes B)=(\det A)^n(\det B)^m,\tag{8}$$

$$\operatorname{tr}(A\otimes B)=(\operatorname{tr}A)(\operatorname{tr}B),\tag{9}$$

$$\operatorname{rank}(A\otimes B)=(\operatorname{rank}A)(\operatorname{rank}B).\tag{10}$$

证明　按定理 7 或推论 8 我们有

$$\det(A\otimes B)=\prod_{i=1}^{m}\prod_{j=1}^{n}(\alpha_i\beta_j)=\Big(\prod_{i=1}^{m}\alpha_i\Big)^n\Big(\prod_{j=1}^{n}\beta_j\Big)^m=(\det A)^n(\det B)^m.$$

$$\operatorname{tr}(A\otimes B)=\sum_{i=1}^{m}\sum_{j=1}^{n}\alpha_i\beta_j=\Big(\sum_{i=1}^{m}\alpha_i\Big)\Big(\sum_{j=1}^{n}\beta_j\Big)=(\operatorname{tr}A)(\operatorname{tr}B).$$

为证(10)式，考虑 A 与 B 的奇异值分解：$A=U\begin{bmatrix}\Sigma & 0\\ 0 & 0\end{bmatrix}V^*$ 与 $B=U_1\begin{bmatrix}\Sigma_1 & 0\\ 0 & 0\end{bmatrix}V_1^*$，则$A\otimes B=(U\otimes U_1)\begin{bmatrix}\Sigma\otimes\Sigma_1 & 0\\ 0 & 0\end{bmatrix}(V\otimes V_1)^*$，因而 $\operatorname{rank}(A\otimes B)=\operatorname{rank}(\Sigma\otimes\Sigma_1)=(\operatorname{rank}A)(\operatorname{rank}B)$. □

两个方阵的张量积往往可以保持原先方阵的一些性质. 下面定理提供这一事

实的例子.

定理 10　设 $A\in \boldsymbol{M}_m(\mathbb{C})$ 与 $B\in \boldsymbol{M}_n(\mathbb{C})$. 若 A 与 B 皆为正规(酉、Hermite、对称、Hermite 正定或 Hermite 正半定)矩阵,则 $A\otimes B\in \boldsymbol{M}_{mn}(\mathbb{C})$ 也是正规(酉、Hermite、对称、Hermite 正定或 Hermite 正半定)矩阵.

本定理的证明留给读者(见习题 2).

我们指出,若 $A,B\in \boldsymbol{M}_n(\mathbb{C})$ 为 Hermite 正定(正半定)矩阵,那么按定理 10, $A\otimes B\in \boldsymbol{M}_{n^2}(\mathbb{C})$ 也为 Hermite 正定(正半定)矩阵. 但是可以证明:A 与 B 的 Hadamard 积 $A\circ B$ 是 $A\otimes B$ 的主子阵(见习题 3),因此,我们顺便地推出 $A\circ B>O(\geqslant O)$. (另一种证明可见第一章 1. 2. 2 小节定理 13)

习题 4. 1. 1

1. 证明推论 4.

2. 证明定理 10.

3. 设 $A,B\in \boldsymbol{M}_n(\mathbb{C})$. 试证:$A\circ B\in \boldsymbol{M}_n(\mathbb{C})$ 为 $A\otimes B$ 的主子阵.

4. 对任意 $A=(a_{ij})\in \boldsymbol{M}_m(\boldsymbol{F})$,试证:

(1) $I_n\otimes A=\mathrm{diag}[A,A,\cdots,A]$.

(2) $I_m\otimes I_n=I_{mn}$.

5. 试证:若 $A\in \boldsymbol{M}_m(\boldsymbol{F})$ 与 $B\in \boldsymbol{M}_n(\boldsymbol{F})$,且 $C=A\oplus B$,则 $e^C=e^B\otimes e^A$;
若 $f(\lambda)$ 为多项式或在 $\sigma(A)$ 上有定义的复值函数,则

$$f(I_n\otimes A)=I_n\otimes f(A),f(A\otimes I_n)=f(A)\otimes I_n.$$

6. 记 $A^{[1]}=A,A^{[k+1]}=A\otimes A^{[k]},k=1,2,\cdots$. 试证:

(1) $A^{[k+l]}=A^{[k]}\otimes A^{[l]}$.

(2) 若 $A,B\in \boldsymbol{M}_n(\boldsymbol{F})$,则 $(AB)^{[k]}=A^{[k]}B^{[k]}$.

7. 设 A,B 为 n 阶 Hermite 正定矩阵. 试证:

(1) $A^{-1}\circ B^{-1}\geqslant (A\circ B)^{-1}$.

(2) $A^{-1}\circ A^{-1}\geqslant (A\circ A)^{-1}$.

(3) $A^{-1}\circ A\geqslant I\geqslant (A^{-1}\circ A)^{-1}$.

4.1.2　矩阵方程的可解条件

设 $A_j\in \boldsymbol{M}_{m,l}(\mathbb{C}),B_j\in \boldsymbol{M}_{n,q}(\mathbb{C}),j=1,2,\cdots,p$,与 $C\in \boldsymbol{M}_{m,q}(\mathbb{C})$ 为给定的矩阵,考虑如下一般线性矩阵方程:

$$A_1XB_1+A_2XB_2+\cdots+A_pXB_p=C, \tag{1}$$

这里 $X\in \boldsymbol{M}_{l,n}(\mathbb{C})$ 为待求的解. 容易看到,通常的线性代数方程组 $A\boldsymbol{x}=\boldsymbol{b}$ 与下列两种方程:

$$AX+XB=C,X-AXB=C \tag{2}$$

均为方程(1)的特殊情形. 特别地,当 $B=A^*$ 时,(2)式中两种方程便是著名的

Ляпунов 方程与 **Stein 方程**. 它们在稳定性理论与控制论以及解析函数插值问题中有着广泛的应用.

本小节应用矩阵的向量函数与张量积工具讨论矩阵方程(1)的可解条件与某些特殊情形的解的表达式.

给定 $A\in \boldsymbol{M}_{m,n}(\mathbb{C})$,用 $A^{(j)}$ 记它的第 j 个列向量,$j=1,\cdots,n$,并将这些列向量依序排成一个$\mathbb{C}^{mn}$中向量:

$$\begin{bmatrix} A^{(1)} \\ A^{(2)} \\ \vdots \\ A^{(n)} \end{bmatrix} \in \mathbb{C}^{mn},$$

我们称之为 A 的**向量函数**,记为 $\mathrm{vec}A$. 它显然由矩阵 A 唯一确定,并且,具有线性性质:

$$\mathrm{vec}(\alpha A+\beta B)=\alpha \mathrm{vec}A+\beta \mathrm{vec}B,\qquad \forall A,B\in \boldsymbol{M}_{mn}(\mathbb{C}) \text{ 与 } \forall \alpha,\beta\in \mathbb{C}.$$

此外,$A_1,A_2,\cdots,A_s$ 在空间 $\boldsymbol{M}_{m,n}(\mathbb{C})$ 中线性无关性等价于 $\mathrm{vec}A_1,\mathrm{vec}A_2,\cdots,\mathrm{vec}A_s$ 在空间$\mathbb{C}^{mn}$中线性无关性.

下面引理 1 给出矩阵向量函数与矩阵张量积之间的一种基本关系.

引理 1　设 $A\in \boldsymbol{M}_{m,l}(\mathbb{C}),B=(b_{ij})\in \boldsymbol{M}_{n,q}(\mathbb{C})$ 与 $X\in \boldsymbol{M}_{l,n}(\mathbb{C})$. 则有

$$\mathrm{vec}(AXB)=(B^{\mathrm{T}}\otimes A)\mathrm{vec}X. \tag{3}$$

证明　$AXB\in \boldsymbol{M}_{m,q}(\mathbb{C})$的第 j 列为

$$(AXB)^{(j)}=AXB^{(j)}=\sum_{k=1}^{n} b_{kj}(AX)^{(k)}=\sum_{k=1}^{n}(b_{kj}A)X^{(k)},$$

因而

$$(AXB)^{(j)}=(b_{1j}A\quad b_{2j}A\quad \cdots\quad b_{nj}A)\mathrm{vec}X,\ j=1,\cdots,q.$$

但$(b_{1j}A\quad b_{2j}A\quad \cdots\quad b_{nj}A)$恰为张量积 $B^{\mathrm{T}}\otimes A$ 的第 j 个块行,于是(3)式得证. □

作为(3)式的直接推论,我们有

$$\mathrm{vec}(AX)=(I_n\otimes A)\mathrm{vec}X\quad (\text{取 } n=q,B=I_n).$$
$$\mathrm{vec}(XB)=(B^{\mathrm{T}}\otimes I_m)\mathrm{vec}X\quad (\text{取 } m=l,A=I_m).$$
$$\mathrm{vec}(AX+XB)=((I_n\otimes A)+(B^{\mathrm{T}}\otimes I_m))\mathrm{vec}X.$$

现将方程(1)写成等价形式:

$$\mathrm{vec}\Big(\sum_{k=1}^{p}A_kXB_k\Big)=\mathrm{vec}C,$$

应用引理 1 可得

$$\sum_{k=1}^{p}(B_k^{\mathrm{T}}\otimes A_k)\mathrm{vec}X=\mathrm{vec}C. \tag{4}$$

令 $G=\sum_{k=1}^{p}(B_k^{\mathrm{T}}\otimes A_k)$,$\boldsymbol{x}=\mathrm{vec}X$ 与 $\boldsymbol{c}=\mathrm{vec}C$,则 $G\in \boldsymbol{M}_{mq,ln}(\mathbb{C})$,$\boldsymbol{x}\in \mathbb{C}^{ln}$ 与 $\boldsymbol{c}\in \mathbb{C}^{mq}$,

且(4) 式可改写为线性代数方程组

$$Gx = c. \tag{5}$$

因此,我们有如下基本结论.

定理 2 线性矩阵方程(1)有解的充分与必要条件为 $c \in \mathrm{Im}G$. 特别地,若各 $A_k \in \boldsymbol{M}_m(\mathbb{C})$, $B_k \in \boldsymbol{M}_n(\mathbb{C})$, $k=1,\cdots,p$,则方程(1)有唯一解的充分与必要条件为 $G \in \boldsymbol{M}_{mn}(\mathbb{C})$非奇异.

在本章我们主要考虑方程(1)的特殊情形(2),并且其中 $A \in \boldsymbol{M}_m(\mathbb{C})$ 与 $B \in \boldsymbol{M}_n(\mathbb{C})$. 由前面分析知道,这时方程(5)分别成为

$$\begin{aligned}((I_n \otimes A) + (B^{\mathrm{T}} \otimes I_m))\mathrm{vec}X &= \mathrm{vec}C, \\ ((I_n \otimes I_m) - (B^{\mathrm{T}} \otimes A))\mathrm{vec}X &= \mathrm{vec}C,\end{aligned} \tag{6}$$

它们有唯一解当且仅当矩阵$(I_n \otimes A)+(B^{\mathrm{T}} \otimes I_m)$与$(I_n \otimes I_m)-(B^{\mathrm{T}} \otimes A)$分别非奇异. 由于 B^{T} 与 B 有相同的特征值,故应用(6)式与 4.1.1 小节的推论 8 便得如下结论.

定理 3 设 $A \in \boldsymbol{M}_m(\mathbb{C})$与 $B \in \boldsymbol{M}_n(\mathbb{C})$. 则当且仅当 $\lambda+\mu \neq 0$, $\forall \lambda \in \sigma(A)$与 $\forall \mu \in \sigma(B)$时,矩阵方程 $AX+XB=C$ 有唯一解;当且仅当 $\lambda\mu \neq 1$, $\forall \lambda \in \sigma(A)$与 $\forall \mu \in \sigma(B)$时,矩阵方程 $X-AXB=C$ 有唯一解.

虽然,上述定理给出矩阵方程(2)有唯一解的充分与必要条件,但要写出它们解矩阵 X 的明显表达式一般是不容易的. 下面基本定理告诉我们,若 $A \in \boldsymbol{M}_m(\mathbb{C})$ 与 $B \in \boldsymbol{M}_n(\mathbb{C})$为**稳定矩阵**,即所有特征值的实部小于零的矩阵,则 $AX+XB=C$ 的解 X 有明显的表达式.

定理 4 设 $A \in \boldsymbol{M}_m(\mathbb{C})$与 $B \in \boldsymbol{M}_n(\mathbb{C})$为稳定矩阵,且 $C \in \boldsymbol{M}_{m,n}(\mathbb{C})$给定. 则线性矩阵方程 $AX+XB=C$ 有唯一解 X,并且,

$$X = -\int_0^{\infty} \mathrm{e}^{At} C \mathrm{e}^{Bt} \,\mathrm{d}t. \tag{7}$$

证明 本定理的假定保证了 A 与 $-B$ 没有相同的特征值,因而按定理 3, $AX+XB=C$ 有唯一解 X. 现考虑初值问题(见第三章习题 3.4.1 题 5):

$$\dot{Z}(t) = AZ(t) + Z(t)B, \qquad Z(0) = C, \tag{8}$$

式中,$Z(t) \in \boldsymbol{M}_{m,n}(\mathbb{C})$为定义在$[0,+\infty)$上的矩阵值函数. 直接验证可得,$Z(t)=\mathrm{e}^{At}C\mathrm{e}^{Bt}$为问题(8)的解. 自 $t=0$ 到 $t=\infty$对方程(8)的两端求积分便有

$$Z(\infty) - Z(0) = A\int_0^{\infty} Z(t)\mathrm{d}t + \left(\int_0^{\infty} Z(t)\mathrm{d}t\right)B,$$

即有

$$A\left(-\int_0^{\infty} \mathrm{e}^{At}C\mathrm{e}^{Bt}\,\mathrm{d}t\right) + \left(-\int_0^{\infty} \mathrm{e}^{At}C\mathrm{e}^{Bt}\,\mathrm{d}t\right)B = C - Z(\infty).$$

因此,为了证明由(7)式确定的 X 为 $AX+XB=C$ 的解,只需验证上式中 $Z(\infty)=\lim\limits_{t\to\infty}\mathrm{e}^{At}C\mathrm{e}^{Bt}=\boldsymbol{O}$. 为此只要证明:$\lim\limits_{t\to\infty}\mathrm{e}^{At}=O$ 对任意稳定矩阵 A 成立. 但按谱分解定

理,我们有

$$e^{At}=\sum_{k=1}^{s}\sum_{j=0}^{m_k-1}t^j e^{\lambda_k t}E_{kj},\qquad t>0,$$

式中,$\lambda_1,\cdots,\lambda_s$ 为 A 不同的特征值,$m_1,\cdots,m_s$ 分别为它们的指标,E_{kj} 为 A 的分量. 现令 $\lambda_k=\alpha_k+i\beta_k$,$\alpha_k$ 与 β_k 分别为 λ_k 的实部与虚部. 按 A 为稳定矩阵的假设,$\alpha_k<0,k=1,2,\cdots,s$,因而

$$|t^j e^{\lambda_k t}|=|t^j e^{\alpha_k t}|\to 0,\qquad t\to\infty.$$

于是,$\lim\limits_{t\to\infty}e^{At}=O$. □

当 $B=A^*$,且 $C=-W$ 时,$AX+XB=C$ 变为著名的 Ляпунов 方程:

$$AX+XA^*=-W.\tag{9}$$

若 A 为 $\boldsymbol{M}_n(\mathbb{C})$ 的稳定矩阵,W 为 Hermite 正半定(正定)矩阵,则定理 4 蕴涵 Ляпунов 方程(9)的唯一解为

$$X=\int_0^{\infty}e^{At}We^{A^*t}\mathrm{d}t.\tag{10}$$

显然,$X=X^*$,且$(X\boldsymbol{y},\boldsymbol{y})=\int_0^{\infty}(We^{A^*t}\boldsymbol{y},e^{A^*t}\boldsymbol{y})\mathrm{d}t\geqslant 0(>0)$,$\forall\,\boldsymbol{y}\neq\boldsymbol{0},\boldsymbol{y}\in\mathbb{C}^n$. 这样一来,我们有如下推论.

推论 5　设 $A\in\boldsymbol{M}_n(\mathbb{C})$ 为稳定矩阵,W 为 Hermite 正半定(正定)矩阵. 则 Ляпунов 方程(9)有唯一解 X,并且它为 Hermite 正半定(正定)矩阵 $X=\int_0^{\infty}e^{At}We^{A^*t}\mathrm{d}t$.

类似地,对于(2)式中第二个方程,若 $A\in\boldsymbol{M}_m(\mathbb{C})$ 与 $B\in\boldsymbol{M}_n(\mathbb{C})$ 满足 $\rho(A)\rho(B)<1$,则按定理 3,$X-AXB=C$ 对任意 $C\in\boldsymbol{M}_{m,n}(\mathbb{C})$ 有唯一解. 直接验证知道,此解有形式:

$$X=\sum_{j=0}^{\infty}A^jCB^j,\tag{11}$$

因为当 $C=O$ 时,(11)式变为 $\boldsymbol{x}=O$,而当 $C\neq O$ 时,按第二章 2.2.2 小节定理 11,$\|A^kCB^k\|^{1/k}\leqslant\|A^k\|^{1/k}\|C\|^{1/k}\|B^k\|^{1/k}\to\rho(A)\rho(B)<1\,(k\to\infty)$,所以应用 Cauchy 判敛准则,级数 $\sum\limits_{j=0}^{\infty}A^jCB^j$ 绝对收敛.

定理 6　设 $A\in\boldsymbol{M}_m(\mathbb{C})$ 与 $B\in\boldsymbol{M}_n(\mathbb{C})$ 满足条件 $\rho(A)\rho(B)<1$,且 $C\in\boldsymbol{M}_{m,n}(\mathbb{C})$ 给定. 则矩阵方程 $X-AXB=C$ 有唯一解

$$X=\sum_{j=0}^{\infty}A^jCB^j.$$

我们指出,当 $B=A^*$ 时,$X-AXB=C$ 变为著名的 Stein 方程:

$$X-AXA^*=C.\tag{12}$$

这时，若 $\rho(A)<1$，则方程(12)唯一解为 $X=\sum_{j=0}^{\infty}A^jC(A^*)^j$. 显然，若 C 为 Hermite 正半定矩阵，则 $X=X^*$，且 $X-C$ 为 Hermite 正半定的. 因此从 $X=C+(X-C)$ 我们有定理 6 的如下推论.

推论 7 设 $A\in \boldsymbol{M}_n(\mathbb{C})$ 满足 $\rho(A)<1$，$C\in \boldsymbol{M}_n(\mathbb{C})$ 为 Hermite 正半定(正定)矩阵. 则 Stein 方程(12) 有唯一解 X，且它为 Hermite 正半定(正定) 矩阵 $X=\sum_{j=0}^{\infty}A^jC(A^*)^j$.

习题 4.1.2

1. 考虑矩阵微分方程

(1) $\dot{X}(t)=AX(t)$;

(2) $\dot{X}(t)=AX(t)+X(t)B$,

其中，$A,B\in \boldsymbol{M}_n(\mathbb{C})$ 为斜 Hermite 矩阵，与 t 无关. 试证：每个方程有一个解 $X(t)$，对所有 t 它为酉矩阵.

2. 设 $A(t),B(t)\in \boldsymbol{M}_n(\mathbb{C})$，$\forall t\in[t_1,t_2]$，且矩阵微分方程 $\dot{X}(t)+A(t)X(t)+X(t)B(t)=O$ 的任意解 $X(t)$ 非奇异，$\forall t\in[t_1,t_2]$. 试证：

(1) 矩阵值函数 $Y(t)=X(t)^{-1}$ 满足线性矩阵方程

$$\dot{Y}(t)-Y(t)A(t)=B(t),\ \forall t\in[t_1,t_2].$$

(2) 若 A 与 t 无关，则(1)中矩阵方程满足条件 $Y(t_0)=C^{-1}(t_1\leqslant t_0\leqslant t_2)$ 的解为

$$Y(t)=C^{-1}\mathrm{e}^{A(t-t_0)}+\int_{t_0}^{t}B(\tau)\mathrm{e}^{A(t-\tau)}\mathrm{d}\tau.$$

3. 给定 $A=\mathrm{diag}(\alpha_1,\cdots,\alpha_n)$(其中诸 α_j 的模小于 1)，$\boldsymbol{J}$ 为元素均是 1 的 n 阶矩阵，而 $\boldsymbol{b}=(b_1,\cdots,b_n)^{\mathrm{T}}$ 与 $\boldsymbol{c}=(c_1,\cdots,c_n)^{\mathrm{T}}\in\mathbb{C}^n$. 证明：矩阵方程 $X-AXA^*=J-\boldsymbol{bc}^*$ 有唯一解. 试求出此解的明显表达式.

4. 设给定 $A=\mathrm{diag}(\alpha_1,\cdots,\alpha_n)$(其中诸 α_j 的虚部>0)，$\boldsymbol{b}=(b_1,\cdots,b_n)^{\mathrm{T}}$ 与 $\boldsymbol{c}=(c_1,\cdots,c_n)^{\mathrm{T}}\in\mathbb{C}^n$. 证明：矩阵方程 $AX-XA^*=\boldsymbol{bc}^*-\boldsymbol{cb}^*$ 有唯一解. 试求出此解的明显表达式.

5. 证明推论 7.

6. 设 $A\in \boldsymbol{M}_m(\mathbb{C})$ 与 $B\in \boldsymbol{M}_n(\mathbb{C})$ 为给定的矩阵. 考虑线性变换 $T_1,T_2\in\mathscr{L}(\boldsymbol{M}_{m,n}(\mathbb{C}))$：

$$T_1(X)=AXB,\quad T_2(X)=AX+XB,\qquad \forall X\in \boldsymbol{M}_{m,n}(\mathbb{C}).$$

试证：T_1 与 T_2 关于 $\boldsymbol{M}_{m,n}(\mathbb{C})$ 的标准基$\{E_{ij}\}(1\leqslant i\leqslant m,1\leqslant j\leqslant n)$(即 $E_{ij}\in \boldsymbol{M}_{m,n}(\mathbb{C})$，它的$(i,j)$位置元素为 1，其余元素为 0)的矩阵表示分别为

$$B^{\mathrm{T}}\otimes A\quad 与\quad (I_n\otimes A)+(B^{\mathrm{T}}\otimes I_m).$$

7. 将 $\boldsymbol{M}_{l,n}(\mathbb{C})$ 视为以 $(X,Y)=\mathrm{tr}(XY^*)$ 为内积的内积空间，并且设变换 $T:\boldsymbol{M}_{l,n}(\mathbb{C})\to \boldsymbol{M}_{m,q}(\mathbb{C})$ 由下式确定：

$$T(X)=\sum_{j=1}^{p}A_jXB_j,\qquad \forall X\in \boldsymbol{M}_{l,n}(\mathbb{C}),$$

式中，各 $A_j\in \boldsymbol{M}_{m,l}(\mathbb{C})$，各 $B_j\in \boldsymbol{M}_{n,q}(\mathbb{C})$. 试证：$T$ 的伴随变换 T^* 由下式确定：

$$T^*(Y)=\sum_{i=1}^{p}A_i^* YB_i^*,\qquad \forall Y\in \boldsymbol{M}_{m,q}(\mathbb{C}).$$

4.1.3　矩阵方程 $AX+XB=C$

设 $A\in\boldsymbol{M}_m(\mathbb{C})$，$B\in\boldsymbol{M}_n(\mathbb{C})$ 与 $C\in\boldsymbol{M}_{m,n}(\mathbb{C})$ 给定. 在 4.1.2 小节中我们讨论了 $AX+XB=C$ 有解与有唯一解的条件. 本小节讨论当此矩阵方程有解时，解的结构形式. 具体步骤如下. 首先取 $B=-A$，$C=O$，我们得到如下特殊的矩阵方程：

$$AX=XA,\qquad A\in\boldsymbol{M}_n(\mathbb{C}).\tag{1}$$

求解方程(1)等价于寻求与 $A\in\boldsymbol{M}_n(\mathbb{C})$ 可交换的所有矩阵. 显然，方程(1)有无限多个解. 事实上，$X=p(A)$（$p(\lambda)$ 为复系数多项式）总是它的解. 我们在下面定理 1 中将给出方程(1)的解的表达式. 接着，应用这个结果讨论 $AX+XB=O$ 解表达式与它的解空间的构造. 由于 $AX+XB=C$ 的解的可以表示为齐次方程 $AX+XB=O$ 的通解与非齐次方程 $AX+XB=C$ 某个特解之和，故最后我们只要讨论 $AX+XB=C$ 有解的条件即可. 我们指出，不同于 4.1.2 小节定理 2 中有解条件，这里的有解条件取另一种形式（见后面定理 6）.

还回到矩阵方程(1). 设 $A=PJP^{-1}$，J 为 A 的 Jordan 正规形式，$Y=P^{-1}XP$. 则方程(1)相当于

$$JY=YJ.\tag{2}$$

现假定 $J=\mathrm{diag}[J_1,J_2,\cdots,J_p]$，这里，$J_s=\lambda_s I_{k_s}+N_{k_s}$ 为对应于特征值 $\lambda_s(1\leqslant s\leqslant p)$ 的 k_s 阶 Jordan 块，N_{k_s} 为 k_s 阶移位幂零 Jordan 块. 将方程(2)中 $Y\in\boldsymbol{M}_n(\mathbb{C})$ 分块使其与 J 的分块相对应，即有 $Y=[Y_{st}]_{s,t=1}^{p}$，$Y_{st}\in\boldsymbol{M}_{k_s,k_t}(\mathbb{C})$. 此时，方程(2)可改写为

$$J_sY_{st}=Y_{st}J_t,\qquad s,t=1,2,\cdots,p$$

或等价地

$$(\lambda_s-\lambda_t)Y_{st}=Y_{st}N_{k_t}-N_{k_s}Y_{st},\qquad s,t=1,2,\cdots,p.\tag{3}$$

考虑两种可能情形.

情形 I　$\lambda_s\neq\lambda_t$. 从(3)式推出，

$$(\lambda_s-\lambda_t)^2Y_{st}=(\lambda_s-\lambda_t)Y_{st}N_{k_t}-N_{k_s}(\lambda_s-\lambda_t)Y_{st}=Y_{st}N_{k_t}^2-2N_{k_s}Y_{st}N_{k_t}+N_{k_s}^2Y_{st}.$$

再用 $\lambda_s-\lambda_t$ 乘上式两边，如此下去可推出，

$$(\lambda_s-\lambda_t)^rY_{st}=\sum_{j=0}^{r}(-1)^j\binom{r}{j}N_{k_s}^jY_{st}N_{k_t}^{r-j},\qquad r=1,2,\cdots,\tag{4}$$

式中，$N_{k_s}^0=I_{k_s}$. 由于 N_{k_s} 为幂零矩阵，故对充分大的 m，$N_{k_s}^m=O$，因而对足够大 r，(4)式右端中 $N_{k_s}^j$ 与 $N_{k_t}^{r-j}$ 至少有一为零矩阵，于是，$(\lambda_s-\lambda_t)^rY_{st}=O$. 按 $\lambda_s\neq\lambda_t$ 的假定，我们有 $Y_{st}=O$.

情形 II　$\lambda_s=\lambda_t$. 此时方程(3)变为

$$Y_{st}N_{k_t}=N_{k_s}Y_{st},\qquad s,t=1,2,\cdots,p.\tag{5}$$

若令 $Y_{st}=(v_{ij})\in \boldsymbol{M}_{k_s,k_t}(\mathbb{C})$，比较上式两端的对应元素，则可列出 k_sk_t 个方程式：

$$v_{i,j-1}=v_{i+1,j},i=1,2,\cdots,k_s;j=1,2,\cdots,k_t. \tag{6}$$

式中约定 $v_{i0}=v_{k_s+1,j}=0$.(6)式表明 Y_{st} 只能取如下三种形式之一：

(a)若 $k_s=k_t$，则 Y_{st} 有形式：

$$Y_{st}=\begin{bmatrix} c_{st}^{(1)} & c_{st}^{(2)} & \cdots & c_{st}^{(k_s)} \\ 0 & c_{st}^{(1)} & \ddots & c_{st}^{(k_s-1)} \\ & & \ddots & \vdots \\ \vdots & & & c_{st}^{(2)} \\ 0 & 0 & \cdots & c_{st}^{(1)} \end{bmatrix}\equiv T_{k_s},$$

式中，$c_{st}^{(i)}\in\mathbb{C},i=1,2,\cdots,k_s$. 这是一个 k_s 阶上三角形 **Toeplitz 矩阵.**

(b)若 $k_s<k_t$，则 $Y_{st}=[\underbrace{O}_{k_t-k_s\text{列}}\quad T_{k_s}]$.

(c)若 $k_s>k_t$，则 $Y_{st}=\begin{bmatrix} T_{k_t} \\ O \end{bmatrix}\}_{k_s-k_t\text{行}}$

综上所述，我们已经证明了下面结果.

定理 1　设 $A\in\boldsymbol{M}_n(\mathbb{C})$ 有形式 $A=PJP^{-1}$，这里 $J=\text{diag}[J_1,J_2,\cdots,J_p]$ 为 A 的 Jordan 正规形式，$J_1,J_2,\cdots,J_p$ 分别是对应于 A 的特征值 $\lambda_1,\lambda_2,\cdots,\lambda_p$ 的 Jordan 块. 则 $X\in\boldsymbol{M}_n(\mathbb{C})$ 为方程(1)解的充分与必要条件为 $X=PYP^{-1}$，其中，$Y=[Y_{st}]$ 是与 J 有相同分块形式的分块矩阵，对 $\lambda_s\neq\lambda_t$，$Y_{st}=O$，对 $\lambda_s\neq\lambda_t$，Y_{st} 取(a)，(b)与(c)三种形式之一.

例 2　设 $A\in\boldsymbol{M}_7(\mathbb{C})$ 为

$$A=P\text{diag}\left[\begin{pmatrix}2&1\\0&2\end{pmatrix},\begin{pmatrix}2&1&0\\0&2&1\\0&0&2\end{pmatrix},\begin{pmatrix}1&1\\0&1\end{pmatrix}\right]P^{-1}.$$

求矩阵方程 $AX=XA$ 的解 X.

显然，$\sigma(A)=\{2,1\}$，$J_1=2I_2+N_2$，$J_2=2I_3+N_3$，$J_3=I_2+N_2$，N_2 与 N_3 为移位幂零矩阵. 应用定理 1，解矩阵 X 有形式：

$$X=P\text{diag}\left[\left(\begin{array}{cc:ccc} v_0 & v_1 & 0 & \alpha_0 & \alpha_1 \\ 0 & v_0 & 0 & 0 & \alpha_0 \\ \hdashline \beta_0 & \beta_1 & \delta_0 & \delta_1 & \delta_2 \\ 0 & \beta_0 & 0 & \delta_0 & \delta_1 \\ 0 & 0 & 0 & 0 & \delta_0 \end{array}\right),\begin{pmatrix}\varepsilon_0&\varepsilon_1\\0&\varepsilon_0\end{pmatrix}\right]P^{-1},$$

其中，$v_0,v_1,\cdots,\varepsilon_1$(共 11 个)可取任意复数.　□

现在我们讨论矩阵方程 $AX=XA$ 解空间 $\mathscr{S}$ 的维数. 显然，$\mathscr{S}$ 为 $\boldsymbol{M}_n(\mathbb{C})$ 的子空

间.

从定理 1 证明看出，方程(1)一般解 X 中任意参数的个数不少于 A 的阶数 n，且只有当 A 的 Jordan 正规形式 J 中各 Jordan 块对应着 A 的不同特征值(即 A 为非减阶矩阵)时，这些任意参数的个数等于 n(即 $Y=P^{-1}XP$ 为 p 个方阵的直和). 一般地，若 α_{st} 表示 A 的初等因子$(\lambda-\lambda_s)^{k_s}$与$(\lambda-\lambda_t)^{k_t}$的最大公因式的次数($1\leqslant s,t\leqslant p$)，亦即：当 $\lambda_s\neq\lambda_t$ 时，$\alpha_{st}=0$；当 $\lambda_s=\lambda_t$ 时，$\alpha_{st}=\min\{k_s,k_t\}$，则按定理 1 的证明，这时 Y_{ss} 有 $k_s=\alpha_{ss}$ 个任意参数，Y_{tt} 有 $k_t=\alpha_{tt}$ 个任意参数，而 Y_{st} 与 Y_{ts} 分别有 α_{st} 个任意参数. 因此，一般解矩阵 X 中共有

$$\alpha=\sum_{s,t=1}^{p}\alpha_{st} \tag{7}$$

个任意参数.

例如，对例 2 中矩阵方程 $AX=XA$ 的一般解矩阵 X 中任意参数个数按公式(7)应为

$$\begin{aligned}\alpha &= \alpha_{11}+\alpha_{22}+\alpha_{33}+\alpha_{12}+\alpha_{21}+\alpha_{13}+\alpha_{31}+\alpha_{23}+\alpha_{32}\\ &= 2+3+2+2+2+0+0+0+0=11.\end{aligned}$$

应用上述结果，我们得出下列推论.

推论 3　设 $A\in\boldsymbol{M}_n(\mathbb{C})$同定理 1. 则矩阵方程(1)的解空间 $\mathscr{S}$的维数 α 由(7)式给出. 并且，若用 $v_1,v_2,\cdots,v_\alpha$ 表示一般解矩阵 X 中的 α 个任意参数，那么，

$$X=\sum_{i=1}^{\alpha}v_iX_i, \tag{8}$$

式中，$X_i\in\boldsymbol{M}_n(\mathbb{C})$表示对应于 $v_i=1,v_j=0(j\neq i)$的方程(1)的解，$i=1,2,\cdots,\alpha$.

接着考虑齐次线性矩阵方程

$$AX+XB=O \tag{9}$$

的解的形式. 主要的技巧是将它归结为前面讨论过的问题. 显然，方程(9)有解. 假定 X 为它的解，则可以验证

$$\begin{bmatrix}I_m & X\\ O & I_n\end{bmatrix}\quad 与\quad \begin{bmatrix}A & O\\ O & -B\end{bmatrix} \tag{10}$$

可交换，反之，若(10)式中两个$(m+n)$阶矩阵可交换，则 X 必定为方程(9)的解. 因此，对(10)式中这两个矩阵应用定理 1 的结果，我们有(见习题 8)

定理 4　设 $A\in\boldsymbol{M}_m(\mathbb{C})$与 $B\in\boldsymbol{M}_n(\mathbb{C})$分别有 Jordan 正规形式

$$\begin{aligned}J_A &= P^{-1}AP=\mathrm{diag}[\lambda_1I_{k_1}+N_{k_1},\cdots,\lambda_pI_{k_p}+N_{k_p}],\\ J_B &= Q^{-1}BQ=\mathrm{diag}[\mu_1I_{r_1}+N_{r_1},\cdots,\mu_qI_{r_q}+N_{r_q}].\end{aligned}$$

则矩阵方程(9)的任意解 X 有形式 $X=PYQ^{-1}$，其中，$Y=[Y_{st}]_{s,t=1}^{p,q}$ ($Y_{st}\in\boldsymbol{M}_{k_s,r_t}(\mathbb{C})$)是矩阵方程 $J_AY+YJ_B=O$ 的一般解，Y 有形式：若 $\lambda_s\neq-\mu_t$，则 $Y_{st}=O$；若$\lambda_s=-\mu_t$，则 Y_{st}取定理 1 中 (a)，(b)与(c)三种形式之一.

类似地，将推论 3 应用到矩阵方程

$$\begin{bmatrix} A & O \\ O & -B \end{bmatrix}\begin{bmatrix} I_m & X \\ O & I_n \end{bmatrix}=\begin{bmatrix} I_m & X \\ O & I_n \end{bmatrix}\begin{bmatrix} A & O \\ O & -B \end{bmatrix}, \tag{11}$$

我们有如下结果(见习题 8).

推论 5 矩阵方程(9)的一般解呈如下形式：

$$X=\sum_{i=1}^{\alpha} v_i X_i, \tag{12}$$

式中，$X_1,\cdots,X_\alpha$ 为方程(9)的线性无关解，而

$$\alpha=\sum_{s=1}^{p}\sum_{t=1}^{q}\alpha_{st}, \tag{13}$$

这里，$\alpha_{st}(1\leqslant s\leqslant p, 1\leqslant t\leqslant q)$ 为 A 的初等因子 $(\lambda-\lambda_s)^{k_s}$ 与 $-B$ 的初等因子 $(\lambda+\mu_t)^{r_t}$ 的最大公因式的次数.

我们顺便指出，(13)式中 α 恰为方程(9)解空间的维数，因而也等于 $\dim(\mathrm{Ker}G)$，其中，$G=(I_n\otimes A)+(B^{\mathrm{T}}\otimes I_m)=A\oplus B^{\mathrm{T}}$.

最后考虑一般非齐次矩阵方程

$$AX+XB=C. \tag{14}$$

它等价于线性代数方程组 $G\boldsymbol{x}=\boldsymbol{c}$，这里，$G=A\oplus B^{\mathrm{T}}$，$\boldsymbol{x}=\mathrm{vec}X$ 与 $\boldsymbol{c}=\mathrm{vec}C$. 因此，若方程(14)可解，则它的解或是唯一的(如 $\sigma(A)\cap\sigma(-B)$ 为空集)，或有无限多个，且其一般解为方程(9)的一般解与方程(14)的某个特解之和. 于是，为了完整地研究方程(14)，余下的只要讨论 $AX+XB=C$ 有解的条件.

下面的基本结果应用非构造性办法给出方程(14)有解的充分与必要条件. 显然，它相当于

$$\mathrm{rank}G=\mathrm{rank}[G\ \vdots\ \boldsymbol{c}].$$

定理 6 设 $A\in\boldsymbol{M}_m(\mathbb{C})$，$B\in\boldsymbol{M}_n(\mathbb{C})$ 与 $C\in\boldsymbol{M}_{m,n}(\mathbb{C})$. 则矩阵方程(14)有解的充分与必要条件为如下两个$(m+n)$阶矩阵相似：

$$\begin{bmatrix} A & O \\ O & -B \end{bmatrix} \text{ 与 } \begin{bmatrix} A & C \\ O & -B \end{bmatrix}. \tag{15}$$

证明 设方程(14)有解 $X\in\boldsymbol{M}_{m,n}(\mathbb{C})$. 则有

$$\begin{bmatrix} I_m & -X \\ O & I_n \end{bmatrix}\begin{bmatrix} A & O \\ O & -B \end{bmatrix}\begin{bmatrix} I_m & X \\ O & I_n \end{bmatrix}=\begin{bmatrix} A & C \\ O & -B \end{bmatrix},$$

$$\begin{bmatrix} I_m & -X \\ O & I_n \end{bmatrix}^{-1}=\begin{bmatrix} I_m & X \\ O & I_n \end{bmatrix},$$

因而(15)式中两个矩阵相似. 反之，假定有非奇异 $S\in\boldsymbol{M}_{m+n}(\mathbb{C})$ 使得

$$\begin{bmatrix} A & O \\ O & -B \end{bmatrix}=S^{-1}\begin{bmatrix} A & C \\ O & -B \end{bmatrix}S. \tag{16}$$

定义变换 $\Omega_1,\Omega_2\in\mathscr{L}(\boldsymbol{M}_{m+n}(\mathbb{C}))$ 如下：

$$\Omega_1(Z)=\begin{bmatrix}A & O\\ O & -B\end{bmatrix}Z-Z\begin{bmatrix}A & O\\ O & -B\end{bmatrix},\qquad \forall Z\in\boldsymbol{M}_{m+n}(\mathbb{C}),$$

$$\Omega_2(Z)=\begin{bmatrix}A & C\\ O & -B\end{bmatrix}Z-Z\begin{bmatrix}A & O\\ O & -B\end{bmatrix},\qquad \forall Z\in\boldsymbol{M}_{m+n}(\mathbb{C}).$$

由直接验证可得出

$$\mathrm{Ker}\Omega_1=\left\{\begin{bmatrix}T & U\\ V & W\end{bmatrix}:AT=TA,AU=-UB,-BV=VA,BW=WB\right\},$$

$$\mathrm{Ker}\Omega_2=\left\{\begin{bmatrix}T & U\\ V & W\end{bmatrix}:AT+CV=TA,AU+CW=-UB,-BV=VA,BW=WB\right\}.$$

容易看出，为证方程(14)有解，只要在 $\mathrm{Ker}\Omega_2$ 中找出一个形如

$$\begin{bmatrix}T & U\\ O & -I_n\end{bmatrix}\tag{17}$$

的 $(m+n)$ 阶矩阵即可，因为假若这样的矩阵存在，则按 $\mathrm{Ker}\Omega_2$ 的结构得出 $AU+C(-I_n)=-UB$，因而 $U\in\boldsymbol{M}_{m,n}(\mathbb{C})$ 便是方程(14)的解.

(16)式蕴涵 $\Omega_1(Z)=S^{-1}\Omega_2(SZ)\ \forall Z\in\boldsymbol{M}_{m+n}(\mathbb{C})$，因而特别有 $\mathrm{Ker}\Omega_2=\{SZ:Z\in\boldsymbol{M}_{m+n}(\mathbb{C})$ 且 $Z\in\mathrm{Ker}\Omega_1\}$. 由于 S 非奇异，这表明

$$\dim(\mathrm{Ker}\Omega_1)=\dim(\mathrm{Ker}\Omega_2).\tag{18}$$

考虑由所有复矩阵对 (V,W) 组成的集合 $\mathscr{R}$，这里，$V\in\boldsymbol{M}_{n,m}(\mathbb{C})$ 与 $W\in\boldsymbol{M}_n(\mathbb{C})$ 满足 $BV+VA=O$ 与 $BW=WB$. 对于通常的数乘与加法运算：

$$\alpha(V,W)=(\alpha V,\alpha W),\qquad \forall\alpha\in\mathbb{C},$$

$$(V_1,W_1)+(V_2,W_2)=(V_1+V_2,W_1+W_2),$$

$\mathscr{R}$ 为一个复线性空间. 定义变换 $\Phi_i\in\mathscr{L}(\mathrm{Ker}\Omega_i,\mathscr{R})(i=1,2)$ 如下：

$$\Phi_i\left(\begin{bmatrix}T & U\\ V & W\end{bmatrix}\right)=(V,W),\qquad i=1,2.$$

这时，

$$\mathrm{Ker}\Phi_1=\mathrm{Ker}\Phi_2=\left\{\begin{bmatrix}T & U\\ O & O\end{bmatrix}:AT=TA,AU=-UB\right\}.\tag{19}$$

另一方面，我们有

$$\mathrm{Im}\Phi_1=\mathrm{Im}\Phi_2=\mathscr{R}.\tag{20}$$

事实上，首先有 $\mathrm{Im}\Phi_1=\mathscr{R}$，因为 $(V,W)\in\mathscr{R}$，则有 $BV+VA=O$ 与 $BW=WB$，这时有

$$\begin{bmatrix}O & O\\ V & W\end{bmatrix}\in\mathrm{Ker}\Omega_1\quad 与\quad \Phi_1\left(\begin{bmatrix}O & O\\ V & W\end{bmatrix}\right)=(V,W),$$

因而 $\mathscr{R} \subset \mathrm{Im}\Phi_1$. 但 $\mathrm{Im}\Phi_1 \subset \mathscr{R}$ 是显然的,因此 $\mathrm{Im}\Phi_1 = \mathscr{R}$,并且,$\mathrm{Im}\Phi_2 \subset \mathrm{Im}\Phi_1$. 根据第一章 1.1.2 小节(11)式,

$$\dim(\mathrm{Ker}\Phi_i) + \dim(\mathrm{Im}\Phi_i) = \dim(\mathrm{Ker}\Omega_i), \qquad i = 1,2,$$

于是,(18)与(19)式蕴涵 $\dim(\mathrm{Im}\Phi_1) = \dim(\mathrm{Im}\Phi_2)$,因而从 $\mathrm{Im}\Phi_2 \subset \mathrm{Im}\Phi_1$ 推出,$\mathrm{Im}\Phi_1 = \mathrm{Im}\Phi_2$.

现考虑 $\mathscr{R}$ 的元素

$$(O, -I_n) = \Phi_1\left(\begin{bmatrix} I_m & O \\ O & -I_n \end{bmatrix}\right), \quad 其中\begin{bmatrix} I_m & O \\ O & -I_n \end{bmatrix} \in \mathrm{Ker}\Omega_1.$$

根据(20)式存在

$$\begin{bmatrix} T_0 & U_0 \\ V_0 & W_0 \end{bmatrix} \in \mathrm{Ker}\Omega_2$$

使得

$$(O, -I_n) = \Phi_2\left(\begin{bmatrix} T_0 & U_0 \\ V_0 & W_0 \end{bmatrix}\right).$$

按变换 Φ_2 的定义,这意味着 $V_0 = O, W_0 = -I_n$,即在 $\mathrm{Ker}\Omega_2$ 中找到一个形如(17)式的矩阵. □

习题 4.1.3

1. 设 $A = SJS^{-1}$ 有第三章习题 3.2.2 题 2 的形式. 试求 $AX = XA$ 的解 X.

2. 试证:$AX = XA$ 的每一个解为 $A \in \boldsymbol{M}_n(\mathbb{C})$ 的多项式函数的充分与必要条件为 A 是非减阶的,即 A 的 Jordan 正规形式中没有两个 Jordan 块对应于相同的特征值.

3. 设 $A \in \boldsymbol{M}_n(\mathbb{C})$ 为简单矩阵,且 $\lambda_1, \lambda_2, \cdots, \lambda_s$ 为其所有不同的特征值. 试证:矩阵方程 $AX = XA$ 的解空间有维数 $\beta_1^2 + \beta_2^2 + \cdots + \beta_s^2$,其中,$\beta_j$ 为 λ_j 的代数重数.

4. 试证:若 $A \in \boldsymbol{M}_n(\mathbb{C})$ 的所有特征值彼此不同,则 $AX = XA$ 的任意解 $X \in \boldsymbol{M}_n(\mathbb{C})$ 为简单矩阵.

5. 试证:简单矩阵 $A, X \in \boldsymbol{M}_n(\mathbb{C})$ 可交换当且仅当它们有相同的 n 个线性无关的特征向量.

6. 设 $A \in \boldsymbol{M}_n(\mathbb{C})$ 正规且所有特征值彼此不同. 证明:$AX = XA$ 的任意解 $X \in \boldsymbol{M}_n(\mathbb{C})$ 为正规的.

7. 设 $A \in \boldsymbol{M}_m(\mathbb{C})$ 与 $B \in \boldsymbol{M}_n(\mathbb{C})$. 证明:$AX = XB$ 有秩为 r 的解 $X \in \boldsymbol{M}_{m,n}(\mathbb{C})$ 当且仅当存在列满秩 $E \in \boldsymbol{M}_{m,r}(\mathbb{C})$ 与行满秩 $X_0 \in \boldsymbol{M}_{r,n}(\mathbb{C})$ 使得 $AE = EJ, X_0B = JX_0$,这里,$J \in \boldsymbol{M}_r(\mathbb{C})$ 取 Jordan 正规形式.

8. 证明定理 4 与推论 5.

9. 试证:$A \in \boldsymbol{M}_n(\mathbb{C})$ 与所有 n 阶复矩阵可变换,当且仅当存在 $\alpha \in \mathbb{C}$ 使得 $A = \alpha I_n$.

4.2 矩阵惯性定理

本节讨论 Ляпунов 与 Stein 稳定性定理以及作为它们的推广的矩阵惯性定理

与一般稳定性定理. 这些有关稳定性的基本结果在微分方程解的稳定性理论中有着重要的应用.

4.2.1 Ляпунов 稳定性定理与 Stein 稳定性定理

我们在 4.1.2 小节中已经引入**稳定矩阵**的概念,并在定理 4 证明里实际上推出,$A\in \boldsymbol{M}_n(\mathbb{C})$为稳定矩阵当且仅当对任给初值 $\boldsymbol{x}_0\in\mathbb{C}^n$,线性系统

$$\dot{\boldsymbol{x}}(t) = A\boldsymbol{x}(t) \tag{1}$$

的解 $\boldsymbol{x}(t)$趋于 $\mathbf{0}$(当 $t\to\infty$时). 此时,称此系统(关于零解)为**渐近稳定的**.

Ляпунов 在 1892 年发现,当给定 Hermite 正定矩阵 $W\in\boldsymbol{M}_n(\mathbb{C})$时,线性系统 $\dot{\boldsymbol{x}}(t)=A\boldsymbol{x}(t)$的渐近稳定性等价于线性矩阵方程

$$AH + HA^* = -W \tag{2}$$

有唯一的 Hermite 正定矩阵解 H. 我们来证明这个结果. 在本章后面我们用 $A\geqslant O$ ($>O$)表示 $A\in\boldsymbol{M}_n(\mathbb{C})$为 Hermite 正半定(正定)矩阵. 类似地,$\leqslant O$($<O$)用到负半定(负定)情形.

定理 1(Ляпунов 稳定性定理)　设 $A,W\in\boldsymbol{M}_n(\mathbb{C})$,且 $W>O$.

(1)若 A 为稳定矩阵,则矩阵方程(2)有唯一解 $H>O$.

(2)若矩阵方程(2)有解 $H>O$,则 A 为稳定矩阵.

证明　(1)的证明见 4.1.2 小节定理 4 后面的说明与推论 5. 现证(2). 设 $\lambda\in\sigma(A^*)$,且 $A^*\boldsymbol{x}=\lambda\boldsymbol{x},\boldsymbol{x}\neq\mathbf{0}$,则 $\boldsymbol{x}^*A=\bar{\lambda}\boldsymbol{x}^*$. 分别用 $\boldsymbol{x}^*$ 与 $\boldsymbol{x}$ 左乘与右乘方程(2)得出

$$(\bar{\lambda}+\lambda)\boldsymbol{x}^*H\boldsymbol{x} = -\boldsymbol{x}^*W\boldsymbol{x} < 0. \tag{3}$$

由于 $\boldsymbol{x}^*H\boldsymbol{x}>0$,故 $\mathrm{Re}\lambda=\frac{1}{2}(\lambda+\bar{\lambda})<0$,因而 A^* 与 A 均为稳定矩阵. □

线性系统 $\dot{\boldsymbol{x}}(t)=A\boldsymbol{x}(t)\ (t\geqslant 0)$的另一个重要概念是系统(关于零解)为**稳定的**,它相当于对任给 $\boldsymbol{x}_0\in\mathbb{C}^n$,系统的满足此初值的解 $\boldsymbol{x}(t)=\mathrm{e}^{At}\boldsymbol{x}_0$ 有界. 显然,它等价于对 $t\geqslant 0$,e^{At}有界. 类似于第二章 2.2.2 小节定理 7 最后结果的证明,我们有如下关于线性系统为稳定的准则.

定理 2　线性系统 $\dot{\boldsymbol{x}}(t)=A\boldsymbol{x}(t)\ (t\geqslant 0)$为稳定的当且仅当

(1)A 的所有特征值的实部小于或等于零;

(2)所有零实部的 A 的特征值对应的 Jordan 块均为一阶的.

类似于定理 1(2)的证明,我们可推出,若有 $H>O$ 使得 $A^*H+HA=-W\leqslant O$,则 A 的所有特征值的实部小于或等于零. 实际上还有如下稍强的结果.

定理 3　设 $A\in\boldsymbol{M}_n(\mathbb{C})$. 若有 $H>O$ 使得 $A^*H+HA=-W\leqslant O$,则线性系统 $\dot{\boldsymbol{x}}(t)=A\boldsymbol{x}(t)\ (t\geqslant 0)$为稳定的.

证明　考虑二次型 $u(\boldsymbol{x})\equiv(H\boldsymbol{x},\boldsymbol{x})=\boldsymbol{x}^*H\boldsymbol{x},\ \forall\boldsymbol{x}\in\mathbb{C}^n$. 因为 $H>O$,所以 $u(\cdot)^{1/2}$为$\mathbb{C}^n$ 上向量范数(椭圆范数)(见第二章习题 2.1.1 题 1),记 $u(\cdot)^{1/2}=$

$\|\cdot\|_H$. 当 $x(t)$ 为方程 $\dot{x}(t)=Ax(t)\,(t\geqslant 0)$ 的解时，对 $t\geqslant 0$ 有

$$\begin{aligned}\frac{\mathrm{d}u(x(t))}{\mathrm{d}t} &= \dot{x}(t)^* Hx(t) + x(t)^* H\dot{x}(t)\\ &= x(t)^*(A^*H+HA)x(t) = -x(t)^* Wx(t) \leqslant 0.\end{aligned}$$

于是，$u(x(t))\leqslant u(x(0))$，这表明：$\|x(t)\|_H\leqslant\|x(0)\|_H\ (\forall t\geqslant 0)$，因而对 $t\geqslant 0$，$x(t)$ 有界. □

Ляпунов 稳定性分析中另一个基本结果是关于系统(1)不稳定且当 $t\to\infty$ 时有趋于无穷大的无界解 $x(t)$ 的一个特征条件.

定理 4　设 $A\in M_n(\mathbb{C})$ 的特征值 $\lambda_1,\lambda_2,\cdots,\lambda_n$ 满足

$$\prod_{i,j=1}^{n}(\lambda_i+\bar{\lambda}_j)\neq 0. \tag{4}$$

则系统(1)有解 $x(t)$ 满足 $\lim\limits_{t\to\infty}\|x(t)\|_2=\infty$ 的充分与必要条件为对任意 $W>O$，线性矩阵方程 $AH+HA^*=-W$ 的解 H 具有负特征值.

应用后面 4.2.2 小节推论 6，容易得到本定理的一个简单证明. 我们指出，按定理 2，更一般地，系统(1)有无界解 $x(t)$ 相当于 A 或有正实部的特征值或 A 有纯虚特征值其对应的 Jordan 块阶数大于 1. 这个结果亦可直接从 e^{At} 的谱分解得到(见习题 7).

对应于连续定常线性系统(1)，我们还可以考虑最简单的离散线性系统

$$x_j = Ax_{j-1},\qquad j=1,2,\cdots, \tag{5}$$

这里，$A\in M_n(\mathbb{C})$，$x_j\in\mathbb{C}^n$. 我们说离散线性系统(5)是**渐近稳定的**，假如对任取 $x_0\in\mathbb{C}^n$，由方程(5)得到的向量列 $\{x_j\}_{j=0}^{\infty}$ 收敛于 $\mathbb{C}^n$ 中零向量；系统(5)是**稳定的**，假如对任取 $x_0\in\mathbb{C}^n$，由方程(5)得到的向量列 $\{x_j\}$ 有界. 根据第二章 2.2.2 小节定理 6 与 7，系统(5)是渐近稳定或稳定的等价于 $\lim\limits_{k\to\infty}A^k=O$ 或 $\{A^k\}$ 有界，它们分别又等价于 $\rho(A)<1$ 或 $\rho(A)\leqslant 1$ 且当 $\rho(A)=1$ 时，所有模为 1 的特征值对应的 Jordan 块阶数为 1. 今后，将满足 $\rho(A)<1$ 的 $A\in M_n(\mathbb{C})$ 叫作**(关于单位圆周)稳定的**. 它对应着离散系统(5)的渐近稳定性.

由定理 1 知道，稳定矩阵 $A\in M_n(\mathbb{C})$ 与 Ляпунов 线性矩阵方程 $AH+HA^*=-W$ 之间有着密切的联系. 从下面讨论看出，(关于单位圆周)稳定矩阵 $A\in M_n(\mathbb{C})$ 则与 Stein 矩阵方程

$$H-AHA^*=V,\qquad V>O \tag{6}$$

之间保持着密切联系. 应用如下的可逆的矩阵变换：

$$C=(A+I)^{-1}(A-I) \tag{7}$$

可将(关于单位圆周)稳定矩阵 $A\in M_n(\mathbb{C})$ 变换为稳定矩阵 C(见第三章 3.2.1 小节例 8(2))，反之亦然，因为变换(7)的逆变换为

$$A=(I-C)^{-1}(I+C). \tag{8}$$

这时，Ляпунов 矩阵方程 $CH+HC^*=-W(W>O)$ 在变换(7)下可写成

$$(A+I)^{-1}[(A-I)H(A^*+I)+(A+I)H(A^*-I)](A^*+I)^{-1}=-W,$$

进而有

$$H-AHA^*=\frac{1}{2}(A+I)W(A^*+I). \tag{9}$$

由于 $W>O$ 与 $\rho(A)<1$，故 $\frac{1}{2}(A+I)W(A^*+I)>O$. 因此，方程(9)为形如(6)式的 Stein 方程. 反之，通过变换(8)，Stein 方程(6)可变换为一个 Ляпунов 方程

$$CH+HC^*=-\frac{1}{2}(C-I)V(C^*-I), \tag{10}$$

其中，$W\equiv\frac{1}{2}(C-I)V(C^*-I)>O$，因为 $V>O$ 且 C 为稳定矩阵.

这样一来，应用前面关于稳定矩阵与 Ляпунов 方程的结论，我们可以得出(关于单位圆周)稳定矩阵与 Stein 方程(6)的类似结论.

定理 5(Stein 稳定性定理)　设 $A,V\in \boldsymbol{M}_n(\mathbb{C})$，且 $V>O$.

(1)若 A 为(关于单位圆周)稳定矩阵，则矩阵方程(6)有唯一解 $H>O$.

(2)若矩阵方程(6)有解 $H>O$，则 A 为(关于单位圆周)稳定矩阵.

综合前面的结果我们可以得到关于**稳定矩阵的基本定理**(它包含 Ляпунов 稳定性与 Stein 稳定性定理).

定理 6　设 $A\in \boldsymbol{M}_n(\mathbb{C})$，且令 $B=(I-A)^{-1}(I+A)$. 则下列诸命题彼此等价：

(1)A 为稳定矩阵.

(2)B 存在，且 B 为(关于单位圆周)稳定矩阵，即 $\rho(B)<1$.

(3)B 存在，且对任意 $V>O$，矩阵方程 $H-BHB^*=V$ 有解 $H>O$.

(4)B 存在，且对任意 $V>O$ 级数 $\sum\limits_{j=0}^{\infty}B^jV(B^*)^j$ 收敛.

(5)B 存在，且级数 $\sum\limits_{j=0}^{\infty}B^j(B^*)^j$ 收敛.

(6)B 存在，且矩阵方程 $H-BHB^*=I$ 有解 $H>O$.

(7)B 存在，且有 $H>O$ 使得 $H-BHB^*>O$.

(8)存在 $H>O$ 使得 $AH+HA^*<O$.

(9)对任意 $W>O$ 矩阵方程 $AH+HA^*=-W$ 有解 $H>O$.

(10)存在 $H>O$ 满足矩阵方程 $AH+HA^*=-I$.

(11)存在 $W>O$ 使得 $WAW^{-1}+W^{-1}A^*W<O$.

(12)存在非奇异矩阵 T 使得 TAT^{-1} 的实部$<O$.

证明　由定理 1 与 5 以及定理 5 前面的说明知道，本定理(1)，(2)，(3) 与(9)彼此等价.(2)⇒(4)：见 4.1.2 小节推论 7.(4)⇒(5) 显然.(5)⇒(6)：取

$H=\sum_{j=0}^{\infty}B^j(B^*)^j$ 即可，它显然是 Hermite 正定矩阵.（6）⇒（7）显然.（7）⇒（8）：见定理 5 前面的说明.（8）⇒（1）：见定理 1（2）（取 $W=-(AH+HA^*)>O$）. 因此，定理 6 前九个命题彼此等价.（9）⇒（10）：显然，取 $W=I$ 即可.（10）⇒（11）：取 $W=H^{-\frac{1}{2}}$，这里，$H^{\frac{1}{2}}$ 为 $H>O$ 的 Hermite 正定的平方根. 则

$$\begin{aligned}WAW^{-1}+W^{-1}A^*W&=W(AW^{-2}+W^{-2}A^*)W\\&=W(AH+HA^*)W=-W^2=-H^{-1}<O.\end{aligned}$$

（11）⇒（12）：显然，因为取 $T=W$，$2\mathrm{Re}(TAT^{-1})=WAW^{-1}+W^{-1}A^*W<O$.（12）⇒（1）：设 $\lambda\in\sigma(A)$ 且 $A\boldsymbol{x}=\lambda\boldsymbol{x}$，$\boldsymbol{x}\neq\boldsymbol{0}$. 令 $\boldsymbol{y}=T\boldsymbol{x}$，则从 $\langle\mathrm{Re}(TAT^{-1})\boldsymbol{y},\boldsymbol{y}\rangle=\frac{1}{2}\langle TA\boldsymbol{x},\boldsymbol{y}\rangle+\frac{1}{2}\langle A^*T^*\boldsymbol{y},\boldsymbol{x}\rangle=\frac{1}{2}(\lambda+\bar{\lambda})\langle\boldsymbol{y},\boldsymbol{y}\rangle<0$ 推出 $\mathrm{Re}\lambda<0$. □

我们指出，本定理中条件（2）～（7）对所有 $B\in\boldsymbol{M}_n(\mathbb{C})$ 是等价的. 事实上，在此定理中，A 为基本的，B 由 A“生成”的. 然而若交换 A 与 B 在定理中的地位，即将 B 视为基本的，令 $A=(B+I)^{-1}(B-I)$，并且定理 6 中关于 A 的命题（即（1），（8）～（12））将都冠以“A 存在且…”，而关于 B 的命题（（2）～（7））去掉“B 存在”，则这样变换后的定理便为**（关于单位圆周）稳定矩阵的基本定理**.

一个值得注意的问题是在不求特征值前提下，如何判定 $A\in\boldsymbol{M}_n(\mathbb{C})$ 为稳定的. 一种办法是利用 4.1 节中知识，直接求解定理 6（10）中方程 $AH+HA^*=-I$，并判别解矩阵 H 的正定性. Carlson 与 Datta[8] 也提出一种较为简便的直接方法. 另一种办法为先求出 A 的特征多项式（如可用克雷洛夫方法——见习题 3），再应用后面 4.3.3 小节中的 Routh-Hurwitz 稳定性定理.

习题 4.2.1

1. 设 $A\in\boldsymbol{M}_n(\mathbb{C})$. 试证：$A$ 为稳定矩阵的充分与必要条件是存在 $P>O$，$Q>O$ 与斜 Hermite 矩阵 S 使得

$$A=P(S-Q).$$

当 A 有此分解时，

$$-v_n\leqslant\mathrm{Re}\lambda\leqslant-v_1,\qquad\forall\lambda\in\sigma(A),$$

式中，v_i 为 PQ 的特征值，$v_1\leqslant v_2\leqslant\cdots\leqslant v_n$.

2. 设 $A\in\boldsymbol{M}_n(\mathbb{C})$ 稳定，且 $P>O$ 使得 $PA+A^*P=-Q<O$，试证：对任意 $Q_1\geqslant O$ 与斜 Hermite 矩阵 S_1，$A+B$ 仍为稳定矩阵，这里，$B=(S_1-Q_1)P$.

3. 设 $A\in\boldsymbol{M}_n(\mathbb{C})$，且 $\boldsymbol{0}\neq\boldsymbol{u}\in\mathbb{C}^n$ 任意. 试证：若 n 个未知数 $x_0,x_1,\cdots,x_{n-1}$ 的线性方程组

$$x_0\boldsymbol{u}+x_1A\boldsymbol{u}+x_2A^2\boldsymbol{u}+\cdots+x_{n-1}A^{n-1}\boldsymbol{u}=-A^n\boldsymbol{u}$$

有唯一解 $a_0,a_1,\cdots,a_{n-1}$，则

$$z^n + a_{n-1}z^{n-1} + \cdots + a_1 z + a_0$$

为 A 的特征多项式 $\det(zI-A)$.

4. 设 $A\in \boldsymbol{M}_n(\mathbb{R})$. 证明下列诸命题彼此等价：

(1)存在正对角矩阵 D 使得 $AD+DA^{\mathrm{T}}>O$.

(2)存在正对角矩阵 E 使得 $B=E^{-1}AE$ 的实部$>O$.

(3)QA 有一个正的对角元素，$\forall$ 非零 $Q\geqslant O$.

5. 试验证：若 $a<0$，则

$$A=\begin{bmatrix} 0 & 1 \\ -1 & a \end{bmatrix}$$

为稳定矩阵. 试证：

$$H=\begin{bmatrix} 4+a^2 & 2a \\ 2a & 4+a^2 \end{bmatrix}>O,$$

且 $AH+HA^*<O$.

6. 试证：若 $\mathrm{Re}A=\dfrac{1}{2}(A+A^*)<O$，则 A 的稳定矩阵，但逆命题不成立.

7. 应用关于 e^{At} 的谱分解表示，证明：若 $A\in \boldsymbol{M}_n(\mathbb{C})$，则 e^{At} 总能表示为 $e^{At}=Z_-(t)+Z_0(t)+Z_+(t)$，$\forall t\geqslant 0$，其中，$\lim\limits_{t\to+\infty} Z_-(t)=O$，$Z_0(t)$ 或是常数矩阵，或当 $t\to+\infty$ 时有界但无极限，而 $Z_+(t)=O$ 或 $Z_+(t)$ 当 $t\to+\infty$ 时无界.

4.2.2 矩阵惯性定理

20 世纪 60 年代初，苏联的 M. G. Krein[11] 与西方的 A. Ostrowski 和 H. Schneider[12]，O. Taussky[10]，D. Carlson[6] 与 H. K. Wimmer[18] 等人分别独立地推广了 Ляпунов 稳定性定理，得到一般惯性定理(见后面定理 5)与有关结果. 本小节主要讨论这些数学家的工作.

首先引入矩阵 $A\in \boldsymbol{M}_n(\mathbb{C})$ 的**惯性**(inertia)的概念并讨论有关的结论.

对一般矩阵 $A\in \boldsymbol{M}_n(\mathbb{C})$，将它的 n 个特征值 $\lambda_1,\lambda_2,\cdots,\lambda_n$ 位于复平面$\mathbb{C}$右半开平面$\{\lambda\in\mathbb{C}:\mathrm{Re}\lambda>0\}$内，位于左半开平面$\{\lambda\in\mathbb{C}:\mathrm{Re}\lambda<0\}$内与位于虚轴$\{\lambda\in\mathbb{C}:\mathrm{Re}\lambda=0\}$上的个数(计入重根)分别记为 $\pi(A)$，$\nu(A)$ 与 $\delta(A)$，则称三重数$\{\pi(A),\nu(A),\delta(A)\}$为 $A\in \boldsymbol{M}_n(\mathbb{C})$ 的**惯性**，记为 $\mathrm{In}A$. 显然，$\pi(A)+\nu(A)+\delta(A)=n$. 若 $\delta(A)=0$，则 A 非奇异；若 A 为 Hermite 矩阵，则 $\pi(A)$，$\nu(A)$ 与 $\delta(A)$ 分别为 A 的正的，负的与零特征值的个数. 因此对 Hermite 矩阵来说，A 非奇异当且仅当 $\delta(A)=0$，此时，$\pi(A)+\nu(A)=\mathrm{rank}A$. 通常，$\pi(A)-\nu(A)$ 叫作 $A\in \boldsymbol{M}_n(\mathbb{C})$ 的**正负号差**(signature)，记为 $\mathrm{sig}(A)$.

我们说 $A,B\in \boldsymbol{M}_n(\mathbb{C})$ 为**相合的**(congruence)，假如有非奇异矩阵 P 使得 $A=PBP^*$. 由于非奇异矩阵 P 为若干个初等变换矩阵之积，因而 A 与 B 相合相当于 A 由对 B 施以行的与对应列的初等变换得到，反之亦然. 显然，两个酉相似矩阵总是相合的，但反之不真. 并且，相合关系也是 $\boldsymbol{M}_n(\mathbb{C})$ 上的等价关系. 此外，若 A 为

Hermite 矩阵,则与 A 相合的矩阵也是 Hermite 矩阵.

类似于相似关系,可以将 $\boldsymbol{M}_n(\mathbb{C})$ 中所有矩阵按相合关系分成若干个等价类,使得每一个类中任意两个矩阵相合,不同类中的矩阵不相合. 对于 Hermite 矩阵 $A\in\boldsymbol{M}_n(\mathbb{C})$,含它的相合关系等价类中有一个形式特别简单的矩阵(见(1)式中的 D_0),称之为 A 的**关于相合的标准型(或法式)**. 确切地我们有

定理 1 Hermite 矩阵 $A\in\boldsymbol{M}_n(\mathbb{C})$ 相合于如下矩阵:

$$D_0=\begin{bmatrix} I_s & O & O \\ O & -I_{r-s} & O \\ O & O & O_{n-r} \end{bmatrix}, \tag{1}$$

式中,$r=\operatorname{rank}A$ 与 $s=\pi(A)$.

证明 应用第一章 1.2.1 小节定理 1,我们有

$$A=UDU^*, \tag{2}$$

式中,$U\in\mathcal{U}_n$;$D=\operatorname{diag}[\lambda_1,\lambda_2,\cdots,\lambda_n]$,$\lambda_j\in\sigma(A)$,且 D 的对角元素前 $s(=\pi(A))$ 个为 A 的正特征值,接着 $\nu(A)$ 个为 A 的负特征值,余下 $n-r$ 个为 A 的零特征值. 这意味着有正对角矩阵 $D_1\equiv\operatorname{diag}[\sqrt{\lambda_1},\cdots,\sqrt{\lambda_s},\sqrt{|\lambda_{s+1}|},\cdots,\sqrt{|\lambda_r|},1,\cdots,1]$ 使得 $A=UD_1D_0D_1U^*$,这里 D_0 由(1)式确定. 按定义,A 相合于 D_0. □

由定理 1 立即看出,在 $\boldsymbol{M}_n(\mathbb{C})$ 中含有 Hermite 矩阵的相合等价类共有 $(n+1)(n+2)/2$ 个,因为 (1)式中 D_0 只有 $(n+1)(n+2)/2$ 种不同的结构. 于是 $\boldsymbol{M}_n(\mathbb{C})$ 中所有 Hermite 矩阵的不同惯性也恰有 $(n+1)(n+2)/2$ 个.

推论 2 设 $A,B\in\boldsymbol{M}_n(\mathbb{C})$ 为 Hermite 矩阵. 则有:

(1) A 为 Hermite 正定的当且仅当对某个非奇异矩阵 P,$A=PP^*$.

(2) A 为 Hermite 正半定的当且仅当对某个矩阵 $P\in\boldsymbol{M}_n(\mathbb{C})$,$A=PP^*$.

(3) A 与 B 相合当且仅当 $\operatorname{In}A=\operatorname{In}B$.

推论 2(3)蕴涵如下著名的 **Sylvester 惯性律.**

推论 3 相合的 Hermite 矩阵有相等的惯性.

下面结果表明可以在比相合较弱的前提下使得两个 Hermite 矩阵有相等的惯性(见习题 10).

定理 4 若 n 阶 Hermite 矩阵 A,B 满足 $\operatorname{rank}A=\operatorname{rank}B$ 与 $A=MBM^*$,这里 $M\in\boldsymbol{M}_n(\mathbb{C})$,则 $\operatorname{In}A=\operatorname{In}B$.

我们指出,定理 4 只假定 $\operatorname{rank}A=\operatorname{rank}B$ 与 $A=MBM^*$,这里 M 不必为非奇异的. 但当 M 非奇异时,$\operatorname{rank}A=\operatorname{rank}B$ 自然成立,此时定理 4 变为 Sylvester 惯性律(推论 3).

现在研究 Krein[11], Ostrowski 与 Schneider[12] 的**一般惯性定理**. 它是 Ляпунов 稳定性定理(4.2.1 小节定理 1)的重要推广. 此种推广在保持 $W>O$ 前提下,给出 $A\in\boldsymbol{M}_n(\mathbb{C})$ 与 Hermite 矩阵 $H\in\boldsymbol{M}_n(\mathbb{C})$ 惯性之间的关系.

定理 5（一般惯性定理）　设 $A\in\boldsymbol{M}_n(\mathbb{C})$.

(1)若 $\delta(A)=0$，则存在 Hermite 矩阵 $H\in\boldsymbol{M}_n(\mathbb{C})$使得

$$AH+HA^*=W \tag{3}$$

为 Hermite 正定矩阵，且有 $\mathrm{In}A=\mathrm{In}H$.

(2)若存在 Hermite 矩阵 H 使得(3)式中矩阵$>O$，则有 $\mathrm{In}A=\mathrm{In}H,\delta(H)=0$，因而特别地有 H 非奇异.

证明　(1)若 $\delta(A)=0$，则 $\mathrm{In}A=\{p,n-p,0\}$，这里 $p\leqslant n$. 这时，A 的 Jordan 正规形式 J 取

$$P^{-1}AP=J=\mathrm{diag}[J_1,J_2], \tag{4}$$

式中，$J_1\in\boldsymbol{M}_p(\mathbb{C}),J_2\in\boldsymbol{M}_{n-p}(\mathbb{C})$，且 $\mathrm{In}J_1=\{p,0,0\}$与 $\mathrm{In}J_2=\{0,n-p,0\}$. 我们寻求满足条件 $\mathrm{In}A=\mathrm{In}H$ 的 Hermite 矩阵 H 使得 $AH+HA^*>O$. 为此，对稳定矩阵$-J_1$ 与 J_2 应用 4.2.1 小节定理 1 便求得 Hermite 负定矩阵 $H_1\in\boldsymbol{M}_p(\mathbb{C})$与 $H_2\in\boldsymbol{M}_{n-p}(\mathbb{C})$使得 $-J_1H_1-H_1J_1^*=I_p$ 与 $J_2H_2+H_2J_2^*=I_{n-p}$. 现令 $H=P\mathrm{diag}[-H_1,H_2]P^*\in\boldsymbol{M}_n(\mathbb{C})$. 显然，$H=H^*$ 且按推论 3，$\mathrm{In}H=\mathrm{In}\mathrm{diag}[-H_1,H_2]=\{p,n-p,0\}$，因而 $\mathrm{In}A=\mathrm{In}H$. 另一方面，我们有

$$\begin{aligned}AH+HA^*&=P\mathrm{diag}[J_1,J_2]P^{-1}P\mathrm{diag}[-H_1,H_2]P^*\\&\quad+P\mathrm{diag}[-H_1,H_2]P^*(P^*)^{-1}\mathrm{diag}[J_1^*,J_2^*]P^*\\&=P\mathrm{diag}[-J_1H_1-H_1J_1^*,J_2H_2+H_2J_2^*]P^*\\&=PP^*>O.\end{aligned}$$

(2)假定 Hermite 矩阵 $H\in\boldsymbol{M}_n(\mathbb{C})$使得(3)式中矩阵 $W>O$. 按照 4.2.1 小节定理 1 的证明及其记号，这时有$(\lambda+\bar{\lambda})\boldsymbol{x}^*H\boldsymbol{x}=\boldsymbol{x}^*W\boldsymbol{x}>O$，于是，$\mathrm{Re}\lambda\neq0,\forall\lambda\in\sigma(A)$，即有 $\delta(A)=0$. 按(1)的推证，我们也有(4)式成立. 现令 $R=P^{-1}H(P^*)^{-1}$，则(3)式可改写为 $JR+RJ^*=W_O\equiv P^{-1}W(P^*)^{-1}>O$. 将 R 与 W_0 按(4)式中 J 的分块形式写成分块矩阵形式，则有

$$\begin{bmatrix}J_1&O\\O&J_2\end{bmatrix}\begin{bmatrix}R_1&R_2\\R_2^*&R_3\end{bmatrix}+\begin{bmatrix}R_1&R_2\\R_2^*&R_3\end{bmatrix}\begin{bmatrix}J_1^*&O\\O&J_2^*\end{bmatrix}=\begin{bmatrix}W_1&W_2\\W_2^*&W_3\end{bmatrix}, \tag{5}$$

因而 $J_1R_1+R_1J_1^*=W_1$ 与 $J_2R_3+R_3J_2^*=W_3$，其中，$W_1\in\boldsymbol{M}_p(\mathbb{C})$与 $W_3\in\boldsymbol{M}_{n-p}(\mathbb{C})$都是 Hermite 正定的，而$-J_1\in\boldsymbol{M}_p(\mathbb{C})$与 $J_2\in\boldsymbol{M}_{n-p}(\mathbb{C})$都是稳定矩阵. 因此 4.2.1 小节定理 1 断言，(5)式中的 $R_1\in\boldsymbol{M}_p(\mathbb{C})$与$-R_3\in\boldsymbol{M}_{n-p}(\mathbb{C})$均为 Hermite 正定矩阵.

按推论 3，$\mathrm{In}H=\mathrm{In}R$. 因此，余下只要证明 $\mathrm{In}R=\mathrm{In}A=\{p,n-p,0\}$即可. 但这个事实直接得自下面的相合关系式(见第一章 1.1.1 小节(28)式)：

$$Q^*RQ=\begin{bmatrix}R_1&O\\O&-R_2^*R_1^{-1}R_2+R_3\end{bmatrix}, \tag{6}$$

其中，

$$Q=\begin{bmatrix} I & -R_1^{-1}R_2 \\ O & I \end{bmatrix}.$$ □

我们指出，在有的文献里，定理 5 改写为如下等价形式.

定理 5′　设 $A\in \boldsymbol{M}_n(\mathbb{C})$. 则当且仅当存在 Hermite 矩阵 H 使得 $AH+HA^*>O$ 时，$\delta(A)=0$. 此时，$\mathrm{In}A=\mathrm{In}H$.

特别地，当 A 为稳定矩阵时，$\mathrm{In}A=\{0,\nu(A),0\}$，定理 5 或定理 5′退化为 4.2.1 小节中 Ляпунов 稳定性定理.

下面的 Taussky 的结果为定理 5 的直接推论. 它是 4.2.1 小节定理 4 的推广，因而它还能给出 4.2.3　小节定理 4 的一种简单的证明(见习题 1).

推论 6　设 $W>O$ 且 $A\in \boldsymbol{M}_n(\mathbb{C})$的特征值 $\lambda_1,\lambda_2,\cdots,\lambda_n$ 满足

$$\prod_{i,j=1}^{n}(\lambda_i+\bar{\lambda}_j)\neq 0. \tag{7}$$

则存在唯一的非奇异 Hermite 矩阵 H，使得 $AH+HA^*=W>O$，且 $\mathrm{In}A=\mathrm{In}H$.

证明　条件(7)蕴涵 $\sigma(A)$与 $\sigma(-A^*)$不相交，且 $\delta(A)=0$，因而应用定理 5 与 4.1.2 小节定理 3，有唯一的 Hermite 矩阵 H 使得 $AH+HA^*=W>O$，且 $\mathrm{In}A=\mathrm{In}H$. 特别地，$\delta(H)=\delta(A)=0$，因此，$H$ 是非奇异的. □

如果在定理 5 中只要求(3)式中矩阵 $AH+HA^*=W\geqslant O$，则可得到定理 5 的 D. Carlson 与 H. Schneider[19]的推广. 为此，先考虑一些预备知识.

假如 $A,B\in \boldsymbol{M}_n(\mathbb{C})$满足条件 $\pi(A)\leqslant\pi(B)$与 $v(A)\leqslant v(B)$(因而还有 $\delta(A)\geqslant\delta(B)$)，则记为 $\mathrm{In}A\leqslant\mathrm{In}B$.

引理 7　设 $A\in \boldsymbol{M}_n(\mathbb{C})$满足 $\delta(A)=0$. 若 Hermite 矩阵 $H\in \boldsymbol{M}_n(\mathbb{C})$使得 $AH+HA^*\geqslant O$，则 $\mathrm{In}H\leqslant\mathrm{In}A$.

证明　按定理 5，有 Hermite 矩阵 $H_0\in \boldsymbol{M}_n(\mathbb{C})$使得 $AH_0+H_0A^*>O$，且 $\mathrm{In}H_0=\mathrm{In}A$. 现令 $H_\varepsilon=H+\varepsilon H_0$，$\forall\varepsilon>0$，则有 $AH_\varepsilon+H_\varepsilon A^*=(AH+HA^*)+\varepsilon(AH_0+H_0A^*)>O$，$\forall\varepsilon>0$. 再次引用定理 5，我们得到

$$\mathrm{In}H_\varepsilon=\mathrm{In}A,\ \forall\varepsilon>0. \tag{8}$$

特别有 $\delta(H_\varepsilon)=\delta(A)=0$，$\forall\varepsilon>0$，因而 A 与 $H_\varepsilon(\varepsilon>0)$都是非奇异的. 这时，从 H_ε 的(实)特征值作为 ε 的函数的连续性(见第六章 6.3.1 小节)推出，当 $\varepsilon\to0$ 时，随着 H_ε 趋于 H，H_ε 的特征值中有一个或数个趋于零(假如 H 奇异)，其他特征值不改变符号. 因此，(8)式蕴涵 $\mathrm{In}H\leqslant\mathrm{In}H_\varepsilon=\mathrm{In}A$. □

引理 8　设 $A\in \boldsymbol{M}_n(\mathbb{C})$. 若非奇异 Hermite 矩阵 $H\in \boldsymbol{M}_n(\mathbb{C})$使得 $AH+HA^*\geqslant O$，则我们有 $\mathrm{In}A\leqslant\mathrm{In}H$.

证明　令 $A_\varepsilon=A+\varepsilon H^{-1}$，$\forall\varepsilon>0$. 则有 $A_\varepsilon H+HA_\varepsilon^*=(AH+HA^*)+2\varepsilon I_n>O$，$\forall\varepsilon>0$. 根据定理 5(2)，$\mathrm{In}A_\varepsilon=\mathrm{In}H$，$\forall\varepsilon>0$，特别有 $\delta(A_\varepsilon)=\delta(H)=0$，$\forall\varepsilon>0$.

这时，A_ε 的特征值作为 ε 的函数的连续性蕴涵 $\mathrm{In}A$ 与 $\mathrm{In}A_\varepsilon$ 之间的差异仅在于当 $\varepsilon\to 0$ 时某些 A_ε 的特征值变为纯虚数(若 $\delta(A)>0$). 因此，$\mathrm{In}H=\mathrm{In}A_\varepsilon\geqslant \mathrm{In}A$. □

综合前述两个引理，我们直接得到定理 5(2)的一种推广.

定理 9(Carlson-Schneider)　设 $A\in\boldsymbol{M}_n(\mathbb{C})$ 满足 $\delta(A)=0$. 若有非奇异 Hermite 矩阵 $H\in\boldsymbol{M}_n(\mathbb{C})$ 使得 $AH+HA^*=W\geqslant O$，则 $\mathrm{In}A=\mathrm{In}H$.

我们指出，在定理 9 中，若 $W>O$，则由定理 5 知道，这时 $\delta(A)=\delta(H)=0$ 成立，因而这时定理 9 退化为定理 5(2). 但是，当 $W\geqslant O$ 时，为了保证 $\mathrm{In}A=\mathrm{In}H$，除了要求 $H=H^*$ 与 $AH+HA^*=W$ 外，一般还要增加条件：$\delta(A)=\delta(H)=0$. 对这个增加条件的一个常用判则涉及矩阵对的可控性(见第三章 3.4.2 小节).

回顾一下，矩阵 $A\in\boldsymbol{M}_n(\mathbb{C})$ 与 $B\in\boldsymbol{M}_{n,m}(\mathbb{C})$ 组成的矩阵对 (A,B) 称为可控的，假如

$$\mathrm{rank}[B\quad AB\quad\cdots\quad A^{n-1}B]=n.\tag{9}$$

定理 10　设 $A\in\boldsymbol{M}_n(\mathbb{C})$ 与 $B\in\boldsymbol{M}_{n,m}(\mathbb{C})$ 组成可控的矩阵对 (A,B). 若有 Hermite 矩阵 $H\in\boldsymbol{M}_n(\mathbb{C})$ 使得 $AH+HA^*=W\geqslant BB^*$，则有 $\delta(A)=\delta(H)=0$ 与 $\mathrm{In}A=\mathrm{In}H$.

证明　根据定理 9，我们只要证明矩阵对 (A,B) 的可控性蕴涵 $\delta(A)=\delta(B)=0$ 即可. 现假定 $\delta(A)\neq 0$. 这时有 $\boldsymbol{0}\neq\boldsymbol{x}\in\mathbb{C}^n$ 使得 $\boldsymbol{x}^*A=\mathrm{i}\alpha\boldsymbol{x}^*$，这里，$\alpha\in\mathbb{R}$，$-\mathrm{i}\alpha\in\sigma(A^*)$，于是，$A^*\boldsymbol{x}=-\mathrm{i}\alpha\boldsymbol{x}$，并且，

$$\boldsymbol{x}^*W\boldsymbol{x}=\boldsymbol{x}^*H\boldsymbol{x}(-\mathrm{i}\alpha+\mathrm{i}\alpha)=0.$$

由于 $W\geqslant BB^*\geqslant O$，上式蕴涵 $0=\boldsymbol{x}^*W\boldsymbol{x}\geqslant\|B^*\boldsymbol{x}\|_2^2$，因而 $\boldsymbol{x}^*B=\boldsymbol{0}^{\mathrm{T}}$. 这时，对于 A^* 的特征向量 $\boldsymbol{x}\neq\boldsymbol{0}$，

$$\boldsymbol{x}^*[B\quad AB\quad\cdots\quad A^{n-1}B]=[\boldsymbol{x}^*B\quad \mathrm{i}\alpha\boldsymbol{x}^*B\cdots(\mathrm{i}\alpha)^{n-1}\boldsymbol{x}^*B]=\boldsymbol{0}^{\mathrm{T}},$$

因而 $\mathrm{rank}[B\quad AB\quad\cdots\quad A^{n-1}B]<n$，此与(9)式矛盾. 因此，$\delta(A)=0$. 现在假定 $\delta(H)\neq 0$ 即 $\det H=0$，于是有 $\boldsymbol{x}\neq\boldsymbol{0}$，使得 $H\boldsymbol{x}=\boldsymbol{0}$，因而 $\boldsymbol{x}^*W\boldsymbol{x}=\boldsymbol{x}^*(AH+HA^*)\boldsymbol{x}=0$，或等价地有 $\boldsymbol{x}^*W^{\frac{1}{2}}=\boldsymbol{0}^{\mathrm{T}}$，此蕴涵 $\boldsymbol{x}^*W=\boldsymbol{0}^{\mathrm{T}}$，进而 $\boldsymbol{x}^*B=\boldsymbol{0}^{\mathrm{T}}$，于是，

$$\boldsymbol{x}^*AH=\boldsymbol{x}^*W-\boldsymbol{x}^*HA^*=\boldsymbol{0}^{\mathrm{T}}.\tag{10}$$

这样一来，

$$\boldsymbol{x}^*ABB^*A^*\boldsymbol{x}\leqslant\boldsymbol{x}^*AWA^*\boldsymbol{x}=\boldsymbol{x}^*(A^2HA^*+AH(A^*)^2)\boldsymbol{x}=0,$$

因而 $\boldsymbol{x}^*AB=\boldsymbol{0}^{\mathrm{T}}$ 与 $\boldsymbol{x}^*AW=\boldsymbol{0}^{\mathrm{T}}$(因为 $W\geqslant O$). 类似于(10)式，

$$\boldsymbol{x}^*A^2H=\boldsymbol{x}^*AW-\boldsymbol{x}^*AHA^*=\boldsymbol{0}^{\mathrm{T}}.$$

因此我们有

$$\boldsymbol{x}^*A^2BB^*(A^*)^2\boldsymbol{x}\leqslant\boldsymbol{x}^*A^2W(A^*)^2\boldsymbol{x}=\boldsymbol{x}^*(A^3H(A^*)^2+A^2H(A^*)^3)\boldsymbol{x}=0.$$

于是，$\boldsymbol{x}^*A^2B=\boldsymbol{0}^{\mathrm{T}}$. 重复这个过程便得 $\boldsymbol{x}^*A^kB=\boldsymbol{0}^{\mathrm{T}}$，$k=0,1,\cdots,n-1$. 这结果表明有 $\boldsymbol{x}\neq O$ 使得

$$x^*[B \quad AB \quad \cdots \quad A^{n-1}B] = \mathbf{0}^T,$$

此也与矩阵对(A,B)可控性矛盾. 因此,$\delta(H)=0$. □

假如在上述定理中,$n=m$ 且 $B=W^{\frac{1}{2}}$(W 的唯一 Hermite 正半定平方根),则有 $AH+HA^*=W=BB^*\geqslant O$,这时,矩阵对$(A,W)$的可控性蕴涵$(A,W^{\frac{1}{2}})=(A,B)$的可控性,这是因为矩阵乘积的秩不超过其任一因子矩阵的秩,按此,

$$\begin{aligned} n &= \operatorname{rank}[W \quad AW \quad \cdots \quad A^{n-1}W] \\ &= \operatorname{rank}([W^{\frac{1}{2}} \quad AW^{\frac{1}{2}} \quad \cdots \quad A^{n-1}W^{\frac{1}{2}}]\operatorname{diag}[W^{\frac{1}{2}},\cdots,W^{\frac{1}{2}}]) \\ &\leqslant \operatorname{rank}[W^{\frac{1}{2}} \quad AW^{\frac{1}{2}} \quad \cdots \quad A^{n-1}W^{\frac{1}{2}}]. \end{aligned}$$

由此可得定理 10 的下面推论.

推论 11(C. T. Chen-Wimmer)　设 $A\in \boldsymbol{M}_n(\mathbb{C})$. 若有 Hermite 矩阵 $H\in \boldsymbol{M}_n(\mathbb{C})$使得 $AH+HA^*=W\geqslant O$,且矩阵对(A,W)可控,则有 $\delta(A)=\delta(H)=0$ 与 $\operatorname{In}A=\operatorname{In}H$.

考虑推论 11 的一个应用例子.

例 12　给定微分方程

$$K\ddot{\boldsymbol{x}}(t)+(L+S)\dot{\boldsymbol{x}}(t)+M\boldsymbol{x}(t)=\mathbf{0}, \tag{11}$$

式中,K,L 与 M 为 n 阶实对称的,S 为实的斜对称矩阵(即有 $S=-S^T$),且 $\delta(K)=0,L\leqslant O$. 显然,$\boldsymbol{x}(t)=e^{\lambda t}\boldsymbol{v}(\boldsymbol{v}\neq\mathbf{0})$为微分方程(11)的解当且仅当

$$F(\lambda)=K\lambda^2+(L+S)\lambda+M \tag{12}$$

为奇异矩阵. 现令 $\boldsymbol{y}(t)=\begin{bmatrix}\boldsymbol{x}(t)\\ \dot{\boldsymbol{x}}(t)\end{bmatrix}$与

$$A=\begin{bmatrix} O & I_n \\ -K^{-1}M & -K^{-1}(L+S)\end{bmatrix}\in \boldsymbol{M}_{2n}(\mathbb{R}).$$

则方程(11)等价于

$$\dot{\boldsymbol{y}}(t)=A\boldsymbol{y}(t).$$

由于 $\det(\lambda I-A^T)=\det(\lambda I-A)=\det K^{-1}\det F(\lambda)$(见第一章 1.1.1 小节(27)式),故 A^T 或 A 的特征值λ 满足 $\det F(\lambda)=0$,反之亦然. 选取

$$H=\begin{bmatrix} M & O\\ O & K\end{bmatrix} \quad 与 \quad W=\begin{bmatrix} O & O\\ O & L\end{bmatrix}.$$

由简单计算可知,$A^TH+HA=-2W\geqslant O$. 应用推论 11,若矩阵对(A^T,W)可控,则 $\operatorname{In}A^T=\operatorname{In}A=\operatorname{In}H=\operatorname{In}M+\operatorname{In}K,\delta(A^T)=\delta(H)=0$. 若 $\pi(H)=0$(因而 $\pi(M)=\pi(K)=0$),则 A^T 与 A 为稳定矩阵. 这表明当(A^T,W)可控时,矩阵 M 与 K 的稳定性确定微分方程系统(11)的渐近稳定性. □

现在讨论 Ляпунов 稳定性定理的另一种推广. 它涉及矩阵对的可控性的概念.

定理 13(Ляпунов 稳定性定理的推广) 设 $A\in \boldsymbol{M}_n(\mathbb{C})$ 与 $B\in \boldsymbol{M}_{n,m}(\mathbb{C})$ 组成的矩阵对 (A,B) 为可控的,且 $W\in \boldsymbol{M}_n(\mathbb{C})$ 满足 $W\geqslant BB^*\geqslant O$. 则 4.2.1 小节定理 1 中(1)与(2)均成立.

证明 (1)若 A 为稳定矩阵,仿照 4.2.1 小节定理 1 的证明,可得 $AH+HA^*=W$ 有唯一解 $H=-\int_0^{\infty}e^{At}We^{A^*t}dt$,它显然为 Hermite 负半定的. 应用题设 $W\geqslant BB^*\geqslant O$, 矩阵对 (A,B) 的可控性以及第三章 3.4.2 小节引理 8 与推论 9,便得 $H\leqslant -\int_0^{\infty}e^{At}BB^*e^{A^*t}dt=-\widetilde{W}(\infty)\leqslant -\widetilde{W}(1)<O$(这里 $\widetilde{W}(t)=\int_0^t e^{A\tau}BB^*e^{A^*\tau}d\tau$), 因而 $H<O$.

(2)它是定理 10 的特殊情形. □

我们指出,当 $A,W\in \boldsymbol{M}_n(\mathbb{C})$ 且 $W>O$ 时,取 $B=W^{\frac{1}{2}}$,则矩阵对 (A,B) 显然为可控的. 因此,定理 13 确实为 Ляпунов 稳定性定理的一种推广.

在前面我们研究了当 $W>O$ 或 $W\geqslant O$ 时线性矩阵方程 $AH+HA^*=W$ 中矩阵 A 与 H 惯性 $\mathrm{In}A$ 与 $\mathrm{In}H$ 之间的关系. 实际上,矩阵 $A\in \boldsymbol{M}_n(\mathbb{C})$ 的惯性 $\mathrm{In}A=\{\pi(A),\nu(A),\delta(A)\}$ 是相对于复平面 $\mathbb{C}$ 的虚轴而言的. 类似地,我们可以考虑矩阵 $A\in \boldsymbol{M}_n(\mathbb{C})$ **关于实轴的惯性**,记为 $\mathrm{In}'A=\{\pi'(A),\nu'(A),\delta'(A)\}$,这里 $\pi'(A)$, $\nu'(A)$ 与 $\delta'(A)$ 分别表示在复平面 $\mathbb{C}$ 实轴上半开平面,下半开平面与实轴上 A 的特征值个数(计入重数). 譬如,若 $A\in \boldsymbol{M}_n(\mathbb{R})$,则有 $\pi'(A)=\nu'(A)$;若 $A=A^*$,则有 $\pi'(A)=\nu'(A)=0$. 显然,一般地有

$$\pi'(A)+\nu'(A)+\delta'(A)=n, \qquad \forall A\in \boldsymbol{M}_n(\mathbb{C}),$$
$$\mathrm{In}(-\mathrm{i}A)=\mathrm{In}'A, \qquad \forall A\in \boldsymbol{M}_n(\mathbb{C}). \tag{13}$$

上述第二个关系式表明,关于虚轴的矩阵惯性问题可以化为关于实轴的矩阵惯性问题,反之亦然.

按照上述思想,我们还可以考虑 $A\in \boldsymbol{M}_n(\mathbb{C})$ **关于单位圆周的惯性**,记为 $\mathrm{In}^{\circ}A=\{\pi^{\circ}(A),\nu^{\circ}(A),\delta^{\circ}(A)\}$,这里,$\pi^{\circ}(A)$,$\nu^{\circ}(A)$ 与 $\delta^{\circ}(A)$ 分别表示在复平面 $\mathbb{C}$ 单位开圆内,单位闭圆外与单位圆周上 A 的特征值个数(计入重数). 应用这个概念,4.2.1 小节(关于单位圆周)稳定矩阵 $A\in \boldsymbol{M}_n(\mathbb{C})$ 等价于 $\mathrm{In}^{\circ}A=\{n,0,0\}$.

应用 4.2.1 小节中变换(8),Stein 方程可变换为一个 Ляпунов 方程,即 $H-AHA^*=V(V>O)$ 变换为 $CH+HC^*=-W(W>O)$,这里,$C=(A+I)^{-1}(A-I)$,$W=\frac{1}{2}(C-I)V(C^*-I)$.

因此,作为一般惯性定理(定理 5)的对应形式与 Stein 稳定性定理(4.2.1 小节定理 5)的推广,我们有

定理 14 设 $A\in \boldsymbol{M}_n(\mathbb{C})$.

(1)若 $\delta^{\circ}(A)=0$,则存在 Hermite 矩阵 $H\in \boldsymbol{M}_n(\mathbb{C})$使得

$$H-AHA^{*}=V \tag{14}$$

为 Hermite 正定矩阵,且有 $\mathrm{In}^{\circ}A=\mathrm{In}H$.

(2)若存在 Hermite 矩阵 H 使得(14)式中矩阵$>O$,则有 $\mathrm{In}^{\circ}A=\mathrm{In}H$, $\delta(H)=0$,因而特别地有 H 非奇异.

证明 若有 Hermite 矩阵 $H\in \boldsymbol{M}_n(\mathbb{C})$使得 $H-AHA^{*}=V>O$,则 A^{*} 的任一特征值 λ 与对应的特征向量 $\boldsymbol{x}(\neq \boldsymbol{0})$必然满足

$$\boldsymbol{x}^{*}V\boldsymbol{x}=\boldsymbol{x}^{*}H\boldsymbol{x}-\boldsymbol{x}^{*}AHA^{*}\boldsymbol{x}=(1-|\lambda|^{2})\boldsymbol{x}^{*}H\boldsymbol{x}>0,$$

因而 $\delta^{\circ}(A)=\delta^{\circ}(A^{*})=0$. 这时由变换 $C=(A+I)^{-1}(A-I)$确定的 $C\in \boldsymbol{M}_n(\mathbb{C})$没有纯虚的特征值,即 $\delta(C)=0$. 此变换的逆变换将 $H-AHA^{*}=V(V>O)$变为矩阵方程 $CH+HC^{*}=-W(W>O)$. 这时有 $\pi^{\circ}(A)=\nu(C)$,$\nu^{\circ}(A)=\pi(C)$. 根据定理5(2),矩阵方程$(-C)H+H(-C)^{*}=W(W>O)$有 $\mathrm{In}(-C)=\mathrm{In}H$,于是,由于 $\mathrm{In}(-C)=\{\nu(C),\pi(C),\delta(C)\}$我们有 $\mathrm{In}^{\circ}A=\mathrm{In}H$,因此(2)成立. 同理,按定理5(1)可得(1). □

定理 14′ 设 $A\in \boldsymbol{M}_n(\mathbb{C})$. 则当且仅当存在 Hermite 矩阵 H 使得 $H-AHA^{*}>O$ 时,$\delta^{\circ}(A)=0$. 此时,$\mathrm{In}^{\circ}A=\mathrm{In}H$.

类似地我们还可以得到关于 Stein 方程的其他对应的结果(见习题 7,8).

习题 4.2.2

1. 试应用推论 6 证明 4.2.1 小节定理 4.

2. 试验证一下如下的 **Schwarz 矩阵** $S\in \boldsymbol{M}_n(\mathbb{C})$:

$$S=\begin{bmatrix} 0 & 1 & 0 & & \cdots & & 0 \\ -c_1 & 0 & 1 & 0 & \cdots & & 0 \\ 0 & -c_2 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ & & & & & & 0 \\ & & & & & 0 & 1 \\ 0 & & \cdots & & 0 & -c_{n-1} & -c_n \end{bmatrix}$$

满足 Ляпунов 方程 $S^{\mathrm{T}}H+HS=W$,这里,

$$W=\mathrm{diag}[0,\cdots,0,-2c_n^2] \text{ 与 } H=\mathrm{diag}[c_1c_2\cdots c_n,c_2c_3,\cdots c_n,\cdots,c_{n-1}c_n,c_n].$$

3. 试证:若题 2 矩阵 S 中各 c_j 均为非零实数,则矩阵对(S^{T},W)可控,且有$\pi(S)=n-k$, $\nu(S)=k$,$\delta(S)=0$,其中,k 等于数列 $c_n,c_{n-1}c_n,c_{n-2}c_{n-1}c_n,\cdots,c_1c_2\cdots c_n$ 中正项的个数.

4. 设 $A\in \boldsymbol{M}_n(\mathbb{C})$,$\mathrm{Re}A>O$,且 $H\in \boldsymbol{M}_n(\mathbb{C})$为 Hermite 矩阵. 试证:$\mathrm{In}(AH)=\mathrm{In}H$.

5. 试证:若 $A\in \boldsymbol{M}_n(\mathbb{C})$满足 $\mathrm{Re}A\leqslant O$ 且矩阵对$(A,\mathrm{Re}A)$可控,则 A 为稳定矩阵.

6. 试证:若 $W_1\geqslant W_2\geqslant O$,则 $\mathrm{Ker}W_1\subset \mathrm{Ker}W_2$.

7. 写出并证明推论 6 关于 Stein 方程的对应结果.

8. 写出并证明定理 9 与定理 10 以及推论 11 的关于 Stein 方程的对应结果.

9. 设 $A\in M_n(\mathbb{R})$，其惯性为 $\mathrm{In}A=\{\pi(A),\nu(A),\delta(A)\}$. 用 $\tau(A)$ 表示 A 在虚轴上特征值的初等因子数. 试证：若 H 为实对称正半定矩阵使得 $AH+HA^{\mathrm{T}}$ 为对称正半定的，则有

$$\mathrm{rank}H\leqslant\pi(A)+\tau(A).$$

10. 证明定理 4.

4.3 Routh-Hurwitz 问题与 Schur-Cohn 问题

由 4.2 节知道，线性微分系统 $\boldsymbol{x}(t)=A\boldsymbol{x}(t)(t\geqslant 0)$ 与离散系统 $\boldsymbol{x}_j=A\boldsymbol{x}_{j-1}(j=1,2,\cdots)$ 的渐近稳定性实际上可分别相当于矩阵 $A\in\boldsymbol{M}_n(\mathbb{C})$ 的特征多项式 $\det(\lambda I-A)$ 的零点都分布在复平面$\mathbb{C}$ 的左半开平面与单位开圆内部. 但由于求多项式零点通常多有困难，自然希望得到一些准则，不必直接求出多项式零点而能判断其零点在复平面上相对于虚轴或相对于单位圆周的分布情况，这类问题便是本节研究的中心课题，即所谓 **Routh-Hurwitz 问题与 Schur-Cohn 问题**.

作为多项式惯性理论的预备知识，除了 4.2 节中介绍过的矩阵惯性理论以外，还需要多项式对的 Bezout 矩阵与结式矩阵的一些结果.

4.3.1 多项式对的 Bezout 矩阵与结式矩阵

设 $a(\lambda)=\sum\limits_{j=0}^{l}a_j\lambda^j$ 与 $b(\lambda)=\sum\limits_{j=0}^{l}b_j\lambda^j$ 是一对复多项式，且 $a(\lambda)$ 与 $b(\lambda)$ 至少有一个为 l 次多项式，我们用 $\mathrm{Bez}(a,b)\equiv(v_{ij})_{i,j=0}^{l-1}\in\boldsymbol{M}_l(\mathbb{C})$ 表示其元素 v_{ij} 由下式生成的矩阵：

$$\sum_{i,j=0}^{l-1}v_{ij}\lambda^i\mu^j=\frac{a(\lambda)b(\mu)-b(\lambda)a(\mu)}{\lambda-\mu}. \tag{1}$$

通常称之为多项式对 $a(\lambda)$，$b(\lambda)$ 的 **Bézout 矩阵**.

例如，多项式对 $a(\lambda)=1-\lambda^2+2\lambda^3$ 与 $b(\lambda)=2-\lambda^2$ 的 Bézout 矩阵 $\mathrm{Bez}(a,b)$ 的元素由下式确定：

$$\begin{aligned}&\frac{1}{\lambda-\mu}[(1-\lambda^2+2\lambda^3)(2-\mu^2)-(2-\lambda^2)(1-\mu^2+2\mu^3)]\\&=0-\mu+4\mu^2-\lambda+4\lambda\mu+0\lambda\mu^2+4\lambda^2+0\lambda^2\mu-2\lambda^2\mu^2,\end{aligned}$$

因此，

$$\mathrm{Bez}(a,b)=\begin{bmatrix}0&-1&4\\-1&4&0\\4&0&-2\end{bmatrix}. \tag{2}$$

与复多项式 $a(\lambda)=\sum\limits_{j=0}^{l}a_j\lambda^j\,(a_l\neq 0)$ 相对应的还有如下 $s\times(l+s)$ 矩阵$[\boldsymbol{a}]^{(s)}$

$(s \geqslant 1)$:

$$[\boldsymbol{a}]^{(s)} \equiv \begin{bmatrix} a_0 & & a_1 & & \cdots a_{l-1} & a_l & & 0 & \cdots & 0 \\ 0 & a_0 & & a_1 & & a_{l-1} & a_l & 0 & & 0 \\ \vdots & \ddots & \ddots & & \ddots & \ddots & \ddots & \ddots & & \vdots \\ & & & & & & & & & 0 \\ 0 & & \cdots & 0 & a_0 & & a_1 & \cdots & a_{l-1} & a_l \end{bmatrix}. \tag{3}$$

若 $a(\lambda)=\sum\limits_{j=0}^{l} a_j\lambda^j\,(\boldsymbol{a}_l \neq 0)$ 与 $b(\lambda)=\sum\limits_{j=0}^{l} a_j\lambda^j$ 为一对复多项式,则如下的 $2l$ 阶复矩阵 $R(a,b)$ 叫作多项式对 $a(\lambda),b(\lambda)$ 的**结式矩阵**(resultant matrix) 或者 ***Sylvester* 矩阵**:

$$R(a,b)=\begin{bmatrix} [\boldsymbol{a}]^{(l)} \\ [\tilde{\boldsymbol{b}}]^{(l)} \end{bmatrix}, \tag{4}$$

其中,若 $b(\lambda)$ 为 $m(<l)$ 次多项式,则 $l\times(2l)$ 矩阵 $[\tilde{\boldsymbol{b}}]^{(l)}$ 中取 $b_{m+1}=\cdots=b_l=0$.

例如,对应于前面例子,多项式对 $a(\lambda),b(\lambda)$ 的结式矩阵为

$$R(a,b)=\begin{bmatrix} 1 & 0 & -1 & 2 & 0 & 0 \\ 0 & 1 & 0 & -1 & 2 & 0 \\ 0 & 0 & 1 & 0 & -1 & 2 \\ 2 & 0 & -1 & 0 & 0 & 0 \\ 0 & 2 & 0 & -1 & 0 & 0 \\ 0 & 0 & 2 & 0 & -1 & 0 \end{bmatrix}. \tag{5}$$

本小节旨在寻求多项式对 $a(\lambda)$ 与 $b(\lambda)$ 的系数与 $\mathrm{Bez}(a,b)$,$R(a,b)$ 之间的关系式,以及 $\det\mathrm{Bez}(a,b)$,$\det R(a,b)$ 与 $a(\lambda),b(\lambda)$ 公共零点之间的联系.

由定义看出,下面命题 1 成立(见习题 1).

命题 1　设 $a(\lambda)=\sum\limits_{j=0}^{l} a_j\lambda^j\,(a_l\neq 0)$ 与 $b(\lambda)=\sum\limits_{j=0}^{l} b_j\lambda^j$ 为一对复多项式. 则有 $\mathrm{Bez}(a,b)=-\mathrm{Bez}(b,a)$ 与 $\mathrm{Bez}(a,b)=\sum\limits_{j=0}^{l} b_j\mathrm{Bez}(a,\lambda^j)$ 并且,$\mathrm{Bez}(a,b)\in \boldsymbol{M}_l(\mathbb{C})$ 为对称矩阵.

现在考虑 $\mathrm{Bez}(a,b)$ 与多项式对 $a(\lambda),b(\lambda)$ 系数之间的关系式. 基本结果出现在后面的命题 2 中. 为此先讨论一些辅助矩阵.

设 $a(\lambda)=\sum\limits_{j=0}^{l} a_j\lambda^j$,下列三角 **Hankel 矩阵**

$$S(a)=\begin{bmatrix} a_1 & a_2 & a_3 & \cdots & a_l \\ a_2 & a_3 & \cdots & a_l & 0 \\ a_3 & & \iddots & \iddots & \\ \vdots & & & & \vdots \\ a_l & 0 & & \cdots & 0 \end{bmatrix} \in \boldsymbol{M}_l(\mathbb{C}) \tag{6}$$

叫作多项式 $a(\lambda)$ 的**对称化子**(symmetrizer). 按定义, $\mathrm{Bez}(a,1)=S(a)$, 且 $S(a)$ 不依赖 a_0, 当 $a_l\neq 0$ 时, $\det S(a)\neq 0$. 多项式 $a(\lambda)_l=\sum_{j=0}^{l} a_j\lambda^j\ (a_l\neq 0)$ 的**(第一)友阵** C_a (它的特征多项式 $\det(\lambda I-C_a)$ 等于 $a(\lambda)/a_l$) 按义为如下 l 阶矩阵:

$$C_a=\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ & & & & \vdots \\ \vdots & \vdots & \vdots & & 0 \\ 0 & 0 & 0 & \cdots & 1 \\ -\tilde{a}_0 & -\tilde{a}_1 & -\tilde{a}_2 & \cdots & -\tilde{a}_{l-1} \end{bmatrix} \in \boldsymbol{M}_l(\mathbb{C}), \tag{7}$$

这里, $\tilde{a}_j=a_j/a_l$, $j=0,1,\cdots,l-1$. 由直接计算可得: 当 $k=1,2,\cdots,l-1$ 时,

$$S(a)C_a^k=\mathrm{diag}[-PS_1P,S_2], \qquad P,S_1\in \boldsymbol{M}_k(\mathbb{C}),S_2\in \boldsymbol{M}_{l-k}(\mathbb{C}). \tag{8}$$

式中, P 为**翻转矩阵**:

$$P=\begin{bmatrix} O & & 1 \\ & \iddots & \\ 1 & & O \end{bmatrix}, \tag{9}$$

S_1 与 S_2 分别表示 $a_1(\lambda)=a_k+a_{k-1}\lambda+\cdots+a_0\lambda^k$ 与 $a_2(\lambda)=a_k+a_{k+1}\lambda+\cdots+a_l\lambda^{l-k}$ 的对称化子 $S(a_1)$ 与 $S(a_2)$. 当 $k=l$ 时,

$$S(a)C_a^l=-PS(\hat{a})P,P\in \boldsymbol{M}_l(\mathbb{C}), \tag{10}$$

式中, P 为 l 阶翻转矩阵, 并且,

$$\hat{a}(\lambda)=\lambda^l a(\lambda^{-1})=a_l+a_{l-1}\lambda+\cdots+a_0\lambda^l, \tag{11}$$

通常, (11) 式中多项式 $\hat{a}(\lambda)$ 称为 $a(\lambda)$ 的**倒易多项式**(reciprocal polynomial). 于是, 前面的 $a_1(\lambda)$ 便为 $a_0+a_1\lambda+\cdots+a_k\lambda^k$ 的倒易多项式. 从前面看到, C_a^k 左乘以 $S(a)$ 后化为对称矩阵, $k=1,\cdots,l$.

对于任意复多项式 $a(\lambda)=\sum_{j=0}^{l}a_j\lambda^j$ 与 $b(\lambda)=\sum_{j=0}^{l}b_j\lambda^j$, 直接验证可得:

$$S(a)PS(b)=S(b)PS(a), \qquad S(\hat{a})PS(\hat{b})=S(\hat{b})PS(\hat{a}). \tag{12}$$

命题 2　设 $a(\lambda)$ 与 $b(\lambda)$ 如同命题 1. 则

(1) $\mathrm{Bez}(a,b)=S(a)S(\hat{b})P-S(b)S(\hat{a})P$.

(2) $\mathrm{Bez}(a,b)=PS(\hat{b})S(a)-PS(\hat{a})S(b)$.

(3)$\mathrm{Bez}(a,b)=S(a)b(C_a)$ ($\mathrm{Bez}(a,b)$的 **Barnett 因式分解**).

(4)

$$\mathrm{Bez}(a,b)=S(a)\begin{bmatrix}\beta^{\mathrm{T}}\\ \beta^{\mathrm{T}}C_a\\ \vdots\\ \beta^{\mathrm{T}}C_a^{l-1}\end{bmatrix},$$

式中,$\boldsymbol{\beta}^{\mathrm{T}}$ 为 $b(C_a)\in \boldsymbol{M}_l(\mathbb{C})$ 的第一行向量$(\tilde{b}_0\ \tilde{b}_1\cdots\ \tilde{b}_{l-1})$,$\tilde{b}_j=b_j-b_la_j/a_l$,$j=0,1,\cdots,l-1$.

证明 (1) 用$\lambda-\mu$乘上(1)式两边得出

$$\sum_{i,j=0}^{l-1}v_{ij}(\lambda^{i+1}\mu^j-\lambda^i\mu^{j+1})=a(\lambda)b(\mu)-b(\lambda)a(\mu).$$

比较上式两边$\lambda^i\mu^j$ 的系数便得

$$v_{i-1,j}-v_{i,j-1}=a_ib_j-a_jb_i,\qquad i,j=0,1,\cdots,l,$$

这里约定 $v_{-1,j}=v_{i,-1}=v_{lj}=v_{il}=0$. 因此,对 $i,j=0,1,\cdots,l-1$,我们有

$$v_{i+p,j-p}-v_{i+p+1,j-p-1}=a_{i+p+1}b_{j-p}-a_{j-p}b_{i+p+1},$$

其中,$p=0,1,\cdots,m$, 而 $m=\min\{j,l-1-i\}$. 相加之则求出

$$v_{ij}=\sum_{p=0}^{m}(a_{i+p+1}b_{j-p}-a_{j-p}b_{i+p+1}).$$

因此,$\mathrm{Bez}(a,b)=(v_{ij})_{i,j=0}^{l-1}\in\boldsymbol{M}_l(\mathbb{C})$有表达式:

$$\mathrm{Bez}(a,b)=\begin{bmatrix}a_1&a_2&a_3&\cdots&a_l\\ a_2&a_3&&\cdots&0\\ a_3&&\cdot^{\cdot^{\cdot}}\ \cdot^{\cdot^{\cdot}}&&\vdots\\ \vdots&&&&\\ a_l&0&&\cdots&0\end{bmatrix}\begin{bmatrix}b_0&b_1&\cdots&b_{l-1}\\ 0&b_0&\cdots&b_{l-2}\\ 0&&\ddots&\vdots\\ \vdots&\ddots&&b_1\\ 0&\cdots&0&b_0\end{bmatrix}$$
$$-\begin{bmatrix}b_1&b_2&\cdots&b_l\\ b_2&b_3&\cdots&0\\ \vdots&\vdots&\cdot^{\cdot^{\cdot}}&\vdots\\ b_l&0&&0\end{bmatrix}\begin{bmatrix}a_0&a_1&\cdots&a_{l-1}\\ 0&a_0&\cdots&a_{l-2}\\ \vdots&&\ddots\ \ddots&\vdots\\ &\ddots&&a_1\\ 0&\cdots&0&a_0\end{bmatrix}$$
$$=S(a)S(\hat{b})P-S(b)S(\hat{a})P. \tag{13}$$

(2) 由(1)并应用 $\mathrm{Bez}(a,b)$与 $S(a)$的对称性即得.

(3)先证 $\mathrm{Bez}(a,\lambda^k)=S(a)C_a^k$,$0\leqslant k\leqslant l$. 取 $b(\lambda)=\lambda^k$,$0\leqslant k\leqslant l$(即有 $b_k=1$,$b_j=0$,$j\neq k$),代入(13)式中第一个等式由(8)与(10)式即得

$$\mathrm{Bez}(a,\lambda^k)=\mathrm{diag}\left[-\begin{bmatrix} O & & & a_0 \\ & & & a_1 \\ & \iddots & \iddots & \vdots \\ a_0 & a_1 & \cdots & a_{k-1}\end{bmatrix},\begin{bmatrix} a_{k+1} & \cdots & a_l \\ \vdots & & \\ & \iddots & \\ a_l & & O\end{bmatrix}\right]$$

$$= S(a)C_a^k, \qquad 1\leqslant k\leqslant l-1,$$

$$\mathrm{Bez}(a,\lambda^l) = S(a)C_a^l.$$

接着应用命题 1 有 $\mathrm{Bez}(a,b)=\sum_{j=0}^{l} b_j \mathrm{Bez}(a,\lambda^j)=S(a)\sum_{j=0}^{l} b_j C_a^j = S(a)b(C_a)$.

(4)按 Cayley-Hamilton 定理,$a(C_a)=O$, 因而,

$$b(C_a) = b(C_a) - a_l^{-1}b_l a(C_a) = \sum_{j=0}^{l-1}\tilde{b}_j C_a^j,$$

式中,$\tilde{b}_j=b_j-a_l^{-1}b_l a_j, j=0,1,\cdots,l-1$. 现令 $\boldsymbol{\beta}^{\mathrm{T}}=(\tilde{b}_0\tilde{b}_1,\cdots\tilde{b}_{l-1})$. 下面只要证:$b(C_a)$的第 i 行为$\boldsymbol{\beta}^{\mathrm{T}}C_a^{i-1}, i=1,2,\cdots,l$. 事实上,记 $b(C_a)$的第 i 行为$\boldsymbol{d}_i^{\mathrm{T}}, i=1,2,\cdots,l$;$\boldsymbol{e}_i$ 为$\mathbb{C}^l$ 的第 i 个单位坐标向量,$i=1,2,\cdots,l$. 则有

$$\boldsymbol{d}_1^{\mathrm{T}} = \boldsymbol{e}_1^{\mathrm{T}}b(C_a) = (\tilde{b}_0\tilde{b}_1 \cdots\tilde{b}_{l-1}) = \boldsymbol{\beta}^{\mathrm{T}},$$

$$\boldsymbol{d}_2^{\mathrm{T}} = \boldsymbol{e}_2^{\mathrm{T}}b(C_a) = \boldsymbol{e}_1^{\mathrm{T}}C_a b(C_a) = \boldsymbol{d}_1^{\mathrm{T}}C_a = \boldsymbol{\beta}^{\mathrm{T}}C_a,$$

$$\vdots$$

$$\boldsymbol{d}_l^{\mathrm{T}} = \boldsymbol{e}_l^{\mathrm{T}}b(C_a) = \boldsymbol{e}_{l-1}^{\mathrm{T}}C_a b(C_a) = \cdots = \boldsymbol{e}_1^{\mathrm{T}}b(C_a)C_a^{l-1} = \boldsymbol{\beta}^{\mathrm{T}}C_a^{l-1},$$

这里用到事实:$\boldsymbol{e}_1^{\mathrm{T}}C_a^j=\boldsymbol{e}_{j+1}^{\mathrm{T}}, j=0,1,\cdots,l-1$. 再引用(3)中的 $\mathrm{Bez}(a,b)$的 Barnett 因式分解便得(4). □

应用 $\mathrm{Bez}(a,b)$的 Barnett 因式分解,我们可以得到多项式对 $a(\lambda)$与 $b(\lambda)$互素的一个等价条件.

定理 3 设多项式对 $a(\lambda),b(\lambda)$ 同命题 1,则当且仅当 $\det \mathrm{Bez}(a,b)\neq 0$ 时,$a(\lambda)$与 $b(\lambda)$没有公共零点.

证明 由于 $a_l\neq 0$ 蕴涵 $S(a)$非奇异,故按 Barnett 因式分解 $\det\mathrm{Bez}(a,b)\neq 0$ 等价于条件 $\det b(C_a)\neq 0$. 现设 $a(\lambda)$ 的零点为 $\lambda_1,\lambda_2,\cdots,\lambda_l$. 则它们应为 $C_a\in \boldsymbol{M}_l(\mathbb{C})$的特征值,因而按谱映射定理,$b(C_a)$的特征值为 $b(\lambda_1),b(\lambda_2),\cdots,b(\lambda_l)$. 由此即得本定理的结果. □

例如,由于(2)式中矩阵 $\mathrm{Bez}(a,b)$非奇异,故 $a(\lambda)=1-\lambda^2+2\lambda^3$ 与 $b(\lambda)=2-\lambda^2$ 互素.

应用多项式对 $a(\lambda),b(\lambda)$的结式矩阵 $R(a,b)$的非奇异性也可以判断 $a(\lambda)$与 $b(\lambda)$互素. 事实上,按前面记号,(4)式中矩阵 $R(a,b)$可以改写为

$$R(a,b) = \begin{bmatrix}[\boldsymbol{a}]^{(l)} \\ [\tilde{\boldsymbol{b}}]^{(l)}\end{bmatrix} = \begin{bmatrix} S(\hat{a})P & PS(a) \\ S(\hat{b})P & PS(b)\end{bmatrix},$$

式中 P 为 l 阶翻转矩阵. 应用命题 2(1)与(12)式,我们有

$$\begin{bmatrix} P & O \\ -S(b) & S(a) \end{bmatrix} R(a,b) = \begin{bmatrix} PS(\hat{a})P & S(a) \\ \mathrm{Bez}(a,b) & O \end{bmatrix} = \begin{bmatrix} I_l & O \\ O & \mathrm{Bez}(a,b) \end{bmatrix} \begin{bmatrix} PS(\hat{a})P & S(a) \\ I_l & O \end{bmatrix}.$$

这表明,$\mathrm{rank}R(a,b) = \mathrm{rank}\ \mathrm{Bez}(a,b) + l$,因而当且仅当 $\mathrm{Bez}(a,b)$ 非奇异时,$R(a,b) \in \boldsymbol{M}_{2l}(\mathbb{C})$ 非奇异. 这时,本命题便是前一定理的直接推论.

我们指出,尽管 $R(a,b)$ 的阶数为 $\mathrm{Bez}(a,b)$ 阶数的一倍,但由于它的元素或为 $a(\lambda)$ 与 $b(\lambda)$ 的系数或为零,故由 $R(a,b)$ 非奇异判定 $a(\lambda)$ 与 $b(\lambda)$ 无公共零点往往比应用 $\mathrm{Bez}(a,b)$ 更为方便. 例如,应用(5)式中 $R(a,b)$ 的非奇异性也能判定 $a(\lambda)=1-\lambda^2+2\lambda^3$ 与 $b(\lambda)=2-\lambda^2$ 互素.

V. Ptak[14] 指出,多项式对 $a(\lambda),b(\lambda)$ 的 Bézout 矩阵与 Stein 型矩阵方程之间有着密切联系,应用它也能导出 $\mathrm{Bez}(a,b)$ 的 Barnett 因式分解(详见习题 11).

习题 4.3.1

1. 证明命题 1.

2. 验证(8),(10)与(12)式.

3. 试证:l 次的多项式对 $a(\lambda)=\sum_{j=0}^{l} a_j\lambda^j, \bar{a}(\lambda)=\sum_{j=0}^{l}\bar{a}_j\lambda^j$ 的 Bézout 矩阵 $\mathrm{Bez}(a,\bar{a})$ 为斜 Hermite 的,且它的元素皆为纯虚数.

4. 试证:多项式对 $a(\lambda)=\sum_{j=0}^{l} a_j\lambda^j\ (a_l\neq 0), b(\lambda)=\sum_{j=0}^{m} a_j\lambda^j\ (b_m\neq 0, m\leqslant l)$ 的 Bézout 矩阵 $B=\mathrm{Bez}(a,b)$ 满足矩阵方程:

$$C_a^{\mathrm{T}}B = BC_a,$$

这里,C_a 由(7)式确定.

5. 设 $\mathrm{Bez}(a,b)$ 同前题. 试证:

$$\mathrm{Bez}(\hat{a},\hat{b}) = -P\mathrm{Bez}(a,b)P,$$

式中,P 由(9)式确定,而 $\hat{a}(\lambda)$ 为 $a(\lambda)$ 的倒易多项式.

6. 设 $a(\lambda)$ 与 $b(\lambda)$ 同题 4. 定义 $a_\alpha=a_\alpha(\lambda)=a(\lambda+\alpha)$ 与 $b_\alpha=b_\alpha(\lambda)=b(\lambda+\alpha)$. 试证:

$$\mathrm{Bez}(a_\alpha,b_\alpha) = V_\alpha^{(l)}\mathrm{Bez}(a,b)(V_\alpha^{(l)})^{\mathrm{T}},$$

式中,$V_\alpha^{(l)} = \left[\binom{j}{i}\alpha^{j-i}\right]_{i,j=0}^{l-1} \in \boldsymbol{M}_l(\mathbb{C})$ 为由 α 确定的 l 阶**广义 Vandermonde 矩阵**,并约定当 $j<i$ 时,$\binom{j}{i}=0$.

7. 设 P 为 l 阶翻转矩阵. 试证:对题 4 中的多项式 $a(\lambda)$ 与 $b(\lambda)$,

$$R^{\mathrm{T}}(-b,a)\begin{bmatrix} P & O \\ O & P \end{bmatrix} R(a,b) = \begin{bmatrix} O & -B \\ B & O \end{bmatrix}, \qquad B=\mathrm{Bez}(a,b).$$

8. 试证：若 $a(\lambda)=\sum_{j=0}^{l}a_j\lambda^j\ (a_l\neq 0)$，$[\boldsymbol{a}]^{(s)}$ 为(3)式中矩阵，且 $\boldsymbol{\lambda}^{(p)}=(1,\lambda,\lambda^2,\cdots,\lambda^{p-1})^{\mathrm{T}}\in\mathbb{C}^p$（$p$ 为正整数），则有

$$[\boldsymbol{a}]^{(s)}\boldsymbol{\lambda}^{(l+s)}=a(\lambda)\boldsymbol{\lambda}^{(s)}.$$

反之，若矩阵 $A\in\boldsymbol{M}_{s,l+s}(\mathbb{C})$ 满足 $A\boldsymbol{\lambda}^{(l+s)}=a(\lambda)\boldsymbol{\lambda}^{(s)}$，则 $A=[\boldsymbol{a}]^{(s)}$.

9. 设多项式对 $a(\lambda),b(\lambda)$ 同题 4. 试证：对任意正整数 s 都有

$$[\boldsymbol{ab}]^{(s)}=[\boldsymbol{a}]^{(s)}[\boldsymbol{b}]^{(l+s)}=[\boldsymbol{b}]^{(s)}[\boldsymbol{a}]^{(m+s)}.$$

10. 设多项式对 $a(\lambda),b(\lambda)$ 同题 4. 试证下列命题彼此等价：

(1) $a(\lambda)$ 与 $b(\lambda)$ 没有公共零点.

(2) 对每个整数 $p\geqslant 0$，$(m+l+2p)\times(m+l+p)$ 矩阵

$$\begin{bmatrix}[\boldsymbol{a}]^{(m+p)}\\ [\boldsymbol{b}]^{(l+p)}\end{bmatrix}$$

的秩为 $m+l+p$.

(3) 有整数 $p\geqslant 0$ 使得

$$\operatorname{rank}\begin{bmatrix}[\boldsymbol{a}]^{(m+p)}\\ [\boldsymbol{b}]^{(l+p)}\end{bmatrix}=m+l+p.$$

(4)

$$\det\begin{bmatrix}[\boldsymbol{a}]^{(m)}\\ [\boldsymbol{b}]^{(l)}\end{bmatrix}\neq 0.$$

(5)

$$\det R(a,b)=\det\begin{bmatrix}[\boldsymbol{a}]^{(l)}\\ [\tilde{\boldsymbol{b}}]^{(l)}\end{bmatrix}\neq 0,$$

式中，$\tilde{b}(\lambda)$ 即为 $b(\lambda)$，但视为不超过 l 次的多项式（例如，$[\tilde{\boldsymbol{b}}]^{(1)}$ 为 $1\times(l+1)$ 矩阵，它的最后 $l-m$ 个元素为零）.

11. 设多项式对 $a(\lambda),b(\lambda)$ 同题 4. 试证：$\mathrm{Bez}(a,b)$ 唯一地满足线性矩阵方程

$$X-SXC_a=\hat{a}(S)Rb(C_a),$$

这里，

$$S=\begin{bmatrix}0&1&&&0\\&0&1&&\\&&&\ddots&\\&&\ddots&&1\\0&&&&0\end{bmatrix}\in\boldsymbol{M}_l(\mathbb{C})\quad 与\quad R=\begin{bmatrix}0&0&\cdots&0\\ \vdots&\vdots&&\vdots\\0&0&\cdots&0\\1&0&\cdots&0\end{bmatrix}\in\boldsymbol{M}_l(\mathbb{C}).$$

并且，应用这个结果推导 $\mathrm{Bez}(a,b)$ 的 Barnett 因式分解.

12. 设 $a(\lambda)$ 与 $\bar{a}(\lambda)$ 如题 3，$B=\mathrm{Bez}(a,\bar{a})$ 非奇异. 试证矩阵方程：

$$C_a^*X-XC_a=e_1e_1^*,\qquad e_1=(1,0,\cdots,0)^{\mathrm{T}}\in\mathbb{C}^l$$

有唯一解 $X=a(C_a^*)^{-1}B\bar{a}(C_a)^{-1}$.

4.3.2 Routh-Hurwitz 问题与 Schur-Cohn 问题：复多项式的情形

给定复系数多项式

$$a(\lambda) = \sum_{j=0}^{l} a_j \lambda^j \qquad (a_l \neq 0), \tag{1}$$

Routh-Hurwitz 问题是研究多项式 $a(\lambda)$ 的零点相对于复平面虚轴的分布，而 **Schur-Cohn 问题**是研究 $a(\lambda)$ 的零点相对于复平面单位圆周的分布. 当 $a(\lambda)/a_l$ 为 $A \in M_l(\mathbb{C})$ 的特征多项式时，前一问题与 A 的(关于虚轴)惯性密切相关，后一问题与 A 的(关于单位圆周)惯性密切相关.

类似于矩阵的惯性，我们用记号 $\mathrm{In}(a)=\{\pi(a),\nu(a),\delta(a)\}$ ($\mathrm{In}'(a))=\{\pi'(a),\nu'(a),\delta'(a)\}$，$\mathrm{In}^\circ(a)=\{\pi^\circ(a),\nu^\circ(a),\delta^\circ(a)\}$ 表示多项式 $a(\lambda)$ 的**惯性(关于实轴、关于单位圆周的惯性)**，这里，$\pi(a)$，$\nu(a)$ 与 $\delta(a)$ 分别表示多项式 $a(\lambda)$ 具有正实部、负实部与纯虚零点的个数(计入重数)；$\pi'(a)$，$\nu'(a)$ 与 $\delta'(a)$ 分别表示多项式 $a(\lambda)$ 具有正虚部、负虚部与实数零点的个数(计入重数)；$\pi^\circ(a)$，$\nu^\circ(a)$，与 $\delta^\circ(a)$ 分别表示 $a(\lambda)$ 的模小于 1、模大于 1 与模等于 1 的零点的个数(计入重数). 特别地，若 $\pi(a)=\delta(a)=0$ ($\nu^\circ(a)=\delta^\circ(a)=0$)，则称多项式 $a(\lambda)$ 是**稳定的(关于单位圆周稳定的)**.

先研究 Routh-Hurwitz 问题. 显然，如令 $\tilde{a}(\lambda)=a(-\mathrm{i}\lambda)$，则有

$$\mathrm{In}(a) = \mathrm{In}'(\tilde{a}), \tag{2}$$

因而求多项式 $a(\lambda)$ 关于实轴的惯性可自然地归结为求(关于虚轴的)惯性问题，反之亦然.

本小节主要结果可溯源到 C. Hermite，1856 年他应用二次型的正负号差给出 $\mathrm{In}'(a)$ 的求值公式. 1926 年，M. Fujiwara 应用 4.3.1 小节中 Bézout 矩阵的知识，重新改写了多年来未受重视的 C. Hermite 结果，给出稳定性分析的系统方法. 此后该问题及其相关的分析问题的研究获得一系列重要的成果.

令 $\bar{a}(\lambda) = \sum_{j=0}^{l} \bar{a}_j \lambda^j$. 考虑(1)式中多项式 $a(\lambda)$ 与 $\bar{a}(\lambda)$ 的 Bézout 矩阵 $B = \mathrm{Bez}(a,\bar{a})$，它是斜 Hermite 的(见习题 4.3.1 题 3)，因而 $\frac{B}{\mathrm{i}}$ 是 Hermite 矩阵. 并且，按 4.3.1 小节定理 3，$\det B \neq 0$ 等价于多项式 $a(\lambda)$ 与 $\bar{a}(\lambda)$ 互素，这表明 $a(\lambda)$ 既无实零点也没有成共轭复数的零点. 因此，特别地，$\det B \neq 0$ 蕴涵 $\delta'(a) = 0$，且 $a(\lambda)$ 不是实多项式.

定理 1(Hermite-Fujiwara) 设多项式对 $a(\lambda) = \sum_{j=0}^{l} a_j \lambda^j$ ($a_l \neq 0$)，$\bar{a}(\lambda) = \sum_{j=0}^{l} \bar{a}_j \lambda^j$ 使得 $\det B = \det \mathrm{Bez}(a,\bar{a}) \neq 0$. 则

$$\delta'(a)=0 \text{ 与 } \mathrm{In}'(a)=\mathrm{In}(\frac{1}{\mathrm{i}}B).$$

证明　设 C_a 为 $a(\lambda)$ 的友阵(见 4.3.1 小节(7)式)，按习题 4.3.1 题 4，$C_a^{\mathrm{T}}B=BC_a$. 但由 4.3.1 小节(7)式看出，

$$C_a^{\mathrm{T}}=C_a^*-[\mathbf{0}\quad \mathbf{0}\quad \cdots\quad \mathbf{0}\quad \boldsymbol{v}],$$

这里，$\boldsymbol{v}=(a_0/a_l-\bar{a}_0/\bar{a}_l,\cdots,a_{l-1}/a_l-\bar{a}_{l-1}/\bar{a}_l)^{\mathrm{T}}$，它满足 $\boldsymbol{v}^*=-\boldsymbol{v}^{\mathrm{T}}$，因此，

$$C_a^*B-BC_a=[\mathbf{0}\quad \mathbf{0}\quad \cdots\quad \mathbf{0}\quad \boldsymbol{v}]B\equiv W.$$

将它改写为

$$(-\mathrm{i}C_a)^*\left(\frac{1}{\mathrm{i}}B\right)+\left(\frac{1}{\mathrm{i}}B\right)(-\mathrm{i}C_a)=W. \tag{3}$$

$B=\mathrm{Bez}(a,\bar{a})$ 非奇异性蕴涵 $\delta'(a)=\delta'(C_a)=0$，因此由 4.2.2 小节(13)式得到 $\delta(-\mathrm{i}C_a)=\delta(\mathrm{i}C_a^*)=0$. 根据 4.2.2 小节定理 9，为证明 $\mathrm{In}'(a)=\mathrm{In}\left(\frac{1}{\mathrm{i}}B\right)$，只要验证 $W\geqslant O$，因为若 $W\geqslant O$，则有 $\mathrm{In}(-\mathrm{i}C_a)^*=\mathrm{In}\left(\frac{1}{\mathrm{i}}B\right)$，但 $\mathrm{In}(-\mathrm{i}C_a)^*=\mathrm{In}(-\mathrm{i}C_a)=\mathrm{In}'(C_a)=\mathrm{In}'(a)$. 应用 $B=\mathrm{Bez}(a,\bar{a})$ 的 Barnett 因式分解：$B=S(a)\bar{a}(C_a)$，我们有

$$\begin{aligned}W&=[\mathbf{0}\quad \mathbf{0}\quad \cdots\quad \mathbf{0}\quad \boldsymbol{v}]B=[\mathbf{0}\quad \mathbf{0}\quad \cdots\quad \mathbf{0}\quad \boldsymbol{v}]S(a)\bar{a}(C_a)\\&=a_l[\boldsymbol{v}\quad \mathbf{0}\quad \cdots\quad \mathbf{0}]\bar{a}(C_a)=a_l\boldsymbol{v}\boldsymbol{e}_1^{\mathrm{T}}\bar{a}(C_a)=|a_l|^2\boldsymbol{v}\boldsymbol{v}^*\geqslant O,\end{aligned}$$

因为

$$\begin{aligned}\boldsymbol{e}_1^{\mathrm{T}}\bar{a}(C_a)&=\sum_{j=0}^{l}\bar{a}_j\boldsymbol{e}_1^{\mathrm{T}}C_a^j=\sum_{j=0}^{l-1}\bar{a}_j\boldsymbol{e}_{j+1}^{\mathrm{T}}+\bar{a}_l\boldsymbol{e}_l^{\mathrm{T}}C_a\\&=-\bar{a}_l\boldsymbol{v}^{\mathrm{T}}=\bar{a}_l\boldsymbol{v}^*.\end{aligned}$$

作为定理 1 的直接推论，我们有

推论 2　设多项式对 $a(\lambda)$，$\bar{a}(\lambda)$ 同定理 1，则当且仅当 $\frac{1}{\mathrm{i}}\mathrm{Bez}(a,\bar{a})>O$ 时，$a(\lambda)$ 的零点均位于复平面实轴的上半开平面内，即 $\pi'(a)=l$.

例 3　考虑复多项式

$$a(\lambda)=\lambda^4+4\mathrm{i}\lambda^3-4\lambda^2-1=(\lambda+\mathrm{i})^2(\lambda^2+2\mathrm{i}\lambda+1).$$

它有零点：$-\mathrm{i},-\mathrm{i},(1\pm\sqrt{2})\mathrm{i}$，即 $\pi'(a)=1,\nu'(a)=3$ 与 $\delta'(a)=0$. 现在用定理 1 验证之. 按 $\mathrm{Bez}(a,\bar{a})$ 的计算公式：

$$\begin{aligned}\mathrm{Bez}(a,\bar{a})&=S(a)S(\hat{\bar{a}})P-S(\bar{a})S(\hat{a})P\\&=\begin{bmatrix}0&-4&4\mathrm{i}&1\\-4&4\mathrm{i}&1&0\\4\mathrm{i}&1&0&0\\1&0&0&0\end{bmatrix}\begin{bmatrix}-1&0&-4&-4\mathrm{i}\\0&-1&0&-4\\0&0&-1&0\\0&0&0&-1\end{bmatrix}\end{aligned}$$

$$-\begin{bmatrix}0 & -4 & -4\mathrm{i} & 1\\ -4 & -4\mathrm{i} & 1 & 0\\ -4\mathrm{i} & 1 & 0 & 0\\ 1 & 0 & 0 & 0\end{bmatrix}\begin{bmatrix}-1 & 0 & -4 & 4\mathrm{i}\\ 0 & -1 & 0 & -4\\ 0 & 0 & -1 & 0\\ 0 & 0 & 0 & -1\end{bmatrix}$$

$$=\mathrm{i}\begin{bmatrix}0 & 0 & -8 & 0\\ 0 & -8 & 0 & 0\\ -8 & 0 & -32 & 0\\ 0 & 0 & 0 & -8\end{bmatrix},$$

因而 $\det\mathrm{Bez}(a,\bar{a})\neq 0$，且$\frac{1}{\mathrm{i}}B$ 的特征值为$-8,-8,-16\pm 8\sqrt{5}$. 根据定理 1，$\mathrm{In}'(a)=\mathrm{In}\left(\frac{1}{\mathrm{i}}B\right)=\{1,3,0\}$. □

应用定理 1 可以得到 Routh-Hurwitz 问题的基本结论(见后面定理 4).

假定多项式对 $a(\lambda),\bar{a}(\lambda)$如前，并用 $\hat{B}$ 记多项式对 $a(\lambda),\bar{a}(-\lambda)$的 Bézout 矩阵. 令 $D=\mathrm{diag}[1,-1,\cdots,(-1)^{l-1}]\in \boldsymbol{M}_l(\mathbb{C})$ 与 $\widetilde{B}=\hat{B}D\in\boldsymbol{M}_l(\mathbb{C})$，通常将这样的 $\widetilde{B}$ 叫作多项式 $a(\lambda)$的 **Routh-Hurwitz 矩阵**. 容易看出，$\det\widetilde{B}\neq 0$(因而 $\det\hat{B}\neq 0$)等价于 $a(\lambda)$与 $\bar{a}(-\lambda)$没有公共零点，或等价地，$a(\lambda)$既没有纯虚数的零点又没有关于虚轴对称的零点. 因此，特别地，$\det\widetilde{B}\neq 0$ 蕴涵 $\delta(a)=0$.

定理 4(Routh-Hurwitz-Fujiwara)　设多项式对 $a(\lambda),\bar{a}(\lambda)$同定理 1，$\widetilde{B}=\hat{B}D$，这里，$\hat{B}$ 为多项式对 $a(\lambda),\bar{a}(-\lambda)$的 Bézout 矩阵，$D=\mathrm{diag}[1,-1,\cdots,(-1)^{l-1}]$. 若 $\det\widetilde{B}\neq 0$，则

$$\pi(a)=\nu(\widetilde{B}),\nu(a)=\pi(\widetilde{B}),\delta(a)=\delta(\widetilde{B})=0.$$

证明　令 $a_1(\lambda)=a(-\mathrm{i}\lambda)$，$\bar{a}_1(\lambda)=\sum_{j=0}^{l}\bar{a}_j\mathrm{i}^j\lambda^j=\bar{a}(\mathrm{i}\lambda)$，$B=\mathrm{Bet}(a_1,\bar{a}_1)$ 与 $F=\mathrm{diag}[\mathrm{i}^{l-1},\mathrm{i}^{l-2},\cdots,\mathrm{i},1]\in\mathscr{U}_l$，直接验证可推得 $F\left(\frac{1}{\mathrm{i}}B\right)F^*=-\widetilde{B}$，于是，$\widetilde{B}$ 是 Hermite 矩阵. 应用定理 1 与 Sylvester 惯性律，可得 $\mathrm{In}\left(\frac{1}{\mathrm{i}}B\right)=\mathrm{In}'(a_1)=\mathrm{In}(a)=\mathrm{In}(-\widetilde{B})$. 因此，

$$\pi'(a_1)=\pi(a)=\pi\left(\frac{1}{\mathrm{i}}B\right)=\pi(-\widetilde{B})=\nu(\widetilde{B}),$$

$$\nu'(a_1)=\nu(a)=\nu\left(\frac{1}{\mathrm{i}}B\right)=\nu(-\widetilde{B})=\pi(\widetilde{B}).$$

同理可证 $\delta(a)=\delta(\widetilde{B})=0$. □

推论 5　设 $a(\lambda),\bar{a}(\lambda)$与 $\widetilde{B}$ 同定理 4，则当且仅当 $\widetilde{B}>O$ 时，复多项式 $a(\lambda)$为稳定的，即 $\nu(a)=l$.

例 6　考虑多项式 $a(\lambda)=\lambda^2-2\lambda+5$. 则 $\bar{a}(-\lambda)=\lambda^2+2\lambda+5$，且对应的 Bézout 矩阵 $\hat{B}$ 为

$$\hat{B}=\begin{bmatrix}-2 & 1\\ 1 & 0\end{bmatrix}\begin{bmatrix}5 & 2\\ 0 & 5\end{bmatrix}-\begin{bmatrix}2 & 1\\ 1 & 0\end{bmatrix}\begin{bmatrix}5 & -2\\ 0 & 5\end{bmatrix}=\begin{bmatrix}-20 & 0\\ 0 & 4\end{bmatrix}.$$

因此，$a(\lambda)$的 Routh-Hurwitz 矩阵 $\widetilde{B}$ 满足

$$\widetilde{B}=\hat{B}D=\begin{bmatrix}-20 & 0\\ 0 & 4\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}=\begin{bmatrix}-20 & 0\\ 0 & -4\end{bmatrix}<O,$$

这时定理 4 断言：$\mathrm{In}(a)=\mathrm{In}(-\widetilde{B})=\{2,0,0\}$. 实际上，$a(\lambda)$的零点为 $1\pm 2\mathrm{i}$.

我们指出，当 $a(\lambda)$为实系数多项式时，由于总有 $\mathrm{Bez}(a,\bar{a})=O$，故定理 1 无法应用. 但对于定理 4，当 $a(\lambda)$为实多项式时，$\hat{B}$ 因而 $\widetilde{B}$ 满足 $j+k$ 为奇数的(j,k)元素均为零. 经过适当的相合变换，可将 $\widetilde{B}$ 变为

$$Q^{\mathrm{T}}\widetilde{B}Q=\mathrm{diag}[B_1,B_2].$$

因此，$\mathrm{In}\widetilde{B}=\mathrm{In}(Q^{\mathrm{T}}\widetilde{B}Q)=\mathrm{In}B_1+\mathrm{In}B_2$，即可由计算两个较低阶矩阵的惯性 $\mathrm{In}B_1$ 与 $\mathrm{In}B_2$ 之和求得 $\mathrm{In}\widetilde{B}$. 例如，实多项式 $a(\lambda)=a_0+a_1\lambda+a_2\lambda^2+a_3\lambda^3+a_4\lambda^4\ (a_4\neq 0)$对应的 $\widetilde{B}$ 取如下形式

$$\widetilde{B}=2\begin{bmatrix}a_0a_1 & 0 & a_0a_3 & 0\\ 0 & a_1a_2-a_0a_3 & 0 & a_1a_4\\ a_0a_3 & 0 & a_2a_3-a_1a_4 & 0\\ 0 & a_1a_4 & 0 & a_3a_4\end{bmatrix},$$

经过相合变换(取 $Q=[\boldsymbol{e}_1\boldsymbol{e}_2\boldsymbol{e}_3\boldsymbol{e}_4]$)，可将 $\widetilde{B}$ 变为

$$Q^{\mathrm{T}}\widetilde{B}Q=2\mathrm{diag}\left[\begin{pmatrix}a_0a_1 & a_0a_3\\ a_0a_3 & a_2a_3-a_1a_4\end{pmatrix},\begin{pmatrix}a_1a_2-a_0a_3 & a_1a_4\\ a_1a_4 & a_3a_4\end{pmatrix}\right].$$

在 4.3.3 小节定理 6 中我们将讨论对于实多项式的 Routh-Hurwitz-Fujiwara 定理的一种变型.

接着研究 Schur-Cohn 问题. 从前面看出，适当地选取与 $a(\lambda)$有关的多项式对的 Bézout 矩阵便可以通过 Ляпунов 方程把多项式 $a(\lambda)$零点分布问题变换为求相应的 Bézout 矩阵(或它与某对角矩阵之积)的惯性计算问题. 我们继续应用这种技巧，可得出 Schur-Cohn 问题的基本结果.

定理 7　设 $a(\lambda)=\sum_{j=0}^{l}a_j\lambda^j\ (a_l\neq 0)$，$\hat{a}(\lambda)\equiv\lambda^l\bar{a}(\lambda^{-1}$ 的倒易多项式)，且 B 为多项式对 $a(\lambda),\hat{a}(\lambda)$ 的 Bézout 矩阵，$\check{B}=PB$，这里 P 为翻转矩阵. 若 $\det\check{B}\neq 0$，则 $\mathrm{In}^{\circ}(a)=\mathrm{In}\check{B}$，$\delta^{\circ}(a)=\delta(\check{B})=0$.

证明　由于 $\det\check{B}\neq 0$ 等价于 $\det B\neq 0$，故按 B 的 Barnett 因式分解有 $\det\hat{a}(C_a)\neq 0$. 同时，容易看出，$\check{B}=\check{B}^*$，这是因为 $\check{B}^*=B^*P=\bar{B}P=(PS(a)S(\bar{a})-P-S(\hat{a})S(\hat{\bar{a}}))P=P(S(a)S(\bar{a})P-S(\hat{a})S(\hat{\bar{a}})P)=PB=\check{B}$(见

4.3.1 小节命题 1 与命题 2(1) 与(2)). 下面证明 Hermite 矩阵 $\check{B} \in \boldsymbol{M}_l(\mathbb{C})$ 满足如下的 Stein 方程：

$$\check{B} - C_a^* \check{B} C_a = (\hat{a}(C_a))^* \boldsymbol{e}_1 \boldsymbol{e}_1^{\mathrm{T}} \hat{a}(C_a), \tag{4}$$

这里，$\boldsymbol{e}_1$ 为$\mathbb{C}^l$ 的第一个单位坐标向量. 但由于 $\det\hat{a}(C_a) \neq 0$，(4)式可以改写为

$$\hat{a}(C_a)^{-*} \check{B} \hat{a}(C_a)^{-1} - C_a^* \hat{a}(C_a)^{-*} \check{B} \hat{a}(C_a)^{-1} C_a = \boldsymbol{e}_1 \boldsymbol{e}_1^{\mathrm{T}}. \tag{5}$$

为验证(5)式成立，用 $\hat{a}(C_a)^*$ 左乘(5)式两端，左端变为

$$\check{B}\hat{a}(C_a)^{-1} - C_a^* \check{B} \hat{a}(C_a)^{-1} C_a = PS(a) - C_a^* PS(a) C_a$$

$$= \begin{bmatrix} a_l & & & \\ a_{l-1} & a_l & & \\ \vdots & \vdots & \ddots & \\ a_1 & a_2 & \cdots & a_l \end{bmatrix} - \begin{bmatrix} 0 & 0 & \cdots & -\bar{a}_0/\bar{a}_l \\ 1 & 0 & \cdots & -\bar{a}_1/\bar{a}_l \\ 0 & 1 & 0\cdots & -\bar{a}_2/\bar{a}_l \\ \vdots & & \ddots \quad \ddots & \vdots \\ 0 & 0 & 0\cdots 1 & -\bar{a}_{l-1}/\bar{a}_l \end{bmatrix} \times$$

$$\begin{bmatrix} 0 & a_l & 0 & \cdots & 0 \\ 0 & a_{l-1} & a_l & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & a_2 & a_3 & \cdots & a_l \\ -a_0 & 0 & 0 & \cdots & 0 \end{bmatrix} = \begin{bmatrix} a_l - a_0\bar{a}_0/\bar{a}_l & 0 & \cdots & 0 \\ a_{l-1} - a_0\bar{a}_1/\bar{a}_l & 0 & \cdots & 0 \\ \vdots & & & \\ a_1 - a_0\bar{a}_{l-1}/\bar{a}_l & 0 & \cdots & 0 \end{bmatrix},$$

而(5)式右端变为 $\hat{a}(C_a)^* \boldsymbol{e}_1\boldsymbol{e}_1^{\mathrm{T}}$，它恰等于上式中最后的矩阵，因为 $\boldsymbol{e}_1^{\mathrm{T}}\hat{a}(C_a) = \sum_{j=0}^{l} \bar{a}_{l-j}\boldsymbol{e}_1^{\mathrm{T}} C_a^j = \sum_{j=0}^{l-1}\bar{a}_{l-j}\boldsymbol{e}_{j+1}^{\mathrm{T}} + \bar{a}_0 \boldsymbol{e}_l^{\mathrm{T}} C_a = \left(\bar{a}_l - \frac{\bar{a}_0 a_0}{a_l}, \cdots, \bar{a}_1 - \frac{\bar{a}_0 a_{l-1}}{a_l}\right)$. 因此，(5)式成立.

从 $\det B \neq 0$ 看出，$a(\lambda)$既没有单位圆周上的零点又没有关于单位圆周对称的零点，因此，特殊地有 $\delta^\circ(a)=0$. 这与(5)式表明方程 $X - C_a^* X C_a = \boldsymbol{e}_1\boldsymbol{e}_1^{\mathrm{T}}$ 有唯一(非奇异的)解 $X = \hat{a}(C_a)^{-*}\check{B}\hat{a}(C_a)^{-1}$. 于是，$\mathrm{In}X = \mathrm{In}\check{B}$. 对 Stein 矩阵方程 $X - C_a^* X C_a = \boldsymbol{e}_1\boldsymbol{e}_1^{\mathrm{T}}$，应用 4.2.2 小节定理 9 的对应结果(见习题 4.2.2 题 8)可得

$$\mathrm{In}X = \mathrm{In}\check{B} = \mathrm{In}^\circ(C_a^*) = \mathrm{In}^\circ(C_a) = \mathrm{In}^\circ(a), \qquad \delta^\circ(a) = \delta(\check{B}) = 0. \qquad \square$$

通常，我们将定理 7 中 Hermite 矩阵 $\check{B} = PB$ 叫作多项式 $a(\lambda)$的 **Schur-Cohn 矩阵**.

推论 8　设 $a(\lambda)$与 $\check{B}$ 同定理 7. 则当且仅当 $\check{B} > O$ 时，多项式 $a(\lambda)$关于单位圆周是稳定的.

例如，对于二次复多项式 $a(\lambda) = \lambda^2 + \mathrm{i}\lambda + 6$，有 $\hat{a}(\lambda) = 6\lambda^2 - \mathrm{i}\lambda + 1$，这时，

$$\check{B} = PB = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\left(\begin{bmatrix} \mathrm{i} & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 1 & -\mathrm{i} \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} -\mathrm{i} & 6 \\ 6 & 0 \end{bmatrix}\begin{bmatrix} 6 & \mathrm{i} \\ 0 & 6 \end{bmatrix}\right)$$

$$=\begin{bmatrix}-35 & -7\mathrm{i}\\ 7\mathrm{i} & -35\end{bmatrix}<O.$$

应用定理 7 有，$\mathrm{In}^{\circ}(a)=\{0,2,0\}$，同时，$\mathrm{In}^{\circ}(\hat{a})=\{2,0,0\}$，因为 $\hat{a}(\lambda)$ 的 Schur-Cohn 矩阵恰为 $-\check{B}$. 实际上 $a(\lambda)$ 的零点为 $2\mathrm{i}$ 与 $-3\mathrm{i}$，均在单位闭圆以外，而 $\hat{a}(\lambda)$ 的零点为 $\frac{\mathrm{i}}{2}$ 与 $-\frac{\mathrm{i}}{3}$，它们均在单位圆内部.

习题 4.3.2

1. 试对例 3 中的复多项式 $a(\lambda)$，求 $\mathrm{Bez}(a,\hat{a})=B$，这时可否应用定理 7 求 $\mathrm{In}^{\circ}(a)$.

2. 设 $a(\lambda)=(\lambda-\frac{\mathrm{i}}{2})^2(\lambda^2+2\mathrm{i}\lambda+1)$. 试用定理 1 验证一下：$\mathrm{In}'(a)=\{3,1,0\}$；试用定理 7 验证一下：$\mathrm{In}^{\circ}(a)=\{2,2,0\}$.

3. 设 $a(\lambda)=\lambda^2-2\lambda+1$. 试用定理 4 验证一下：$\mathrm{In}(a)=\{2,0,0\}$.

4. 设 $a(\lambda)=(\lambda+1)^2(\lambda^2+\lambda+2)$. 试用定理 4 与例 6 后面的说明验证一下：$\mathrm{In}(a)=\{0,4,0\}$，即 $a(\lambda)$ 为稳定的多项式.

4.3.3 Routh-Hurwitz 问题：实多项式的情形

当 $a(\lambda)=\sum_{j=0}^{l}a_j\lambda^j\,(a_l\neq 0)$ 为实系数多项式时，在求它的零点相对于虚轴分布的 Routh-Hurwitz 问题中，有许多著名的结果，其中包括 Liénard-Chipart 稳定性准则，Markov 准则与便于计算 $\mathrm{In}(\boldsymbol{a})$ 的 Routh-Hurwitz 定理的一种变型（见后面定理 13）.

首先介绍 Lancaster 与 Tismenetsky[13] 关于实多项式扰动的一个定理，它将 $a(\lambda)$ 带有扰动 $\mathrm{i}b(\lambda)=\mathrm{i}\sum_{j=0}^{l-1}b_j\lambda^j$（其中 $b(\lambda)$ 为实多项式）的扰动多项式 $\tilde{a}(\lambda)=a(\lambda)+\mathrm{i}b(\lambda)$（关于实轴）的惯性 $\mathrm{In}'(\tilde{a})$ 与 $\mathrm{Bez}(a,b)$ 的惯性 $\mathrm{InBez}(a,b)$ 联系起来. 应用它可得到实多项式稳定性的准则.

定理 1　设 $a(\lambda)=\sum_{j=0}^{l}a_j\lambda^j\,(a_l\neq 0)$ 与 $b(\lambda)=\sum_{j=0}^{l-1}b_j\lambda^j$ 均为实多项式，$\tilde{a}(\lambda)=a(\lambda)+\mathrm{i}b(\lambda)$. 若 $B=\mathrm{Bez}(a,b)$ 非奇异，则

$$\delta'(\tilde{a})=\delta(B)=0, \tag{1}$$

$$\pi'(a)\leqslant\pi'(\tilde{a})=v(B),\qquad v'(a)\leqslant v'(\tilde{a})=\pi(B), \tag{2}$$

并且，假如 $\delta'(a)=0$，则(2)式中等式均成立.

证明　先来证明 $\delta'(\tilde{a})=0$. 事实上. 若 $\delta'(\tilde{a})\neq 0$，则有 $\lambda_0\in\mathbb{R}$ 使得 $\tilde{a}(\lambda_0)=0$，因而 $a(\lambda_0)=b(\lambda_0)=0$. 但此与实对称矩阵 $B=\mathrm{Bez}(a,b)$ 非奇异性（即有 $\delta(B)=0$）矛盾（4.3.1 小节定理 3）. 接着证明：$\mathrm{In}'(\tilde{a})=\mathrm{In}(-B)$.

为此令 $\tilde{a}(\lambda)=a(\lambda)-\mathrm{i}b(\lambda)$，则 $\mathrm{Bze}(\tilde{a},\bar{\tilde{a}})=\mathrm{Bez}(a+\mathrm{i}b,a-\mathrm{i}b)=-2\mathrm{i}\mathrm{Bez}(a,b)$，因而$\frac{1}{\mathrm{i}}\mathrm{Bez}(\tilde{a},\bar{\tilde{a}})=-2\mathrm{Bez}(a,b)=-2B$ 是非奇异的实对称矩阵. 对 $\tilde{a}(\lambda)$应用 4.3.2 小节定理 1 得到，$\mathrm{In}'(\tilde{a})=\mathrm{In}\left(\frac{1}{\mathrm{i}}\mathrm{Bez}(\tilde{a},\bar{\tilde{a}})\right)=\mathrm{In}(-B)$.

因此为证(2)式成立，我们只需证明：$\pi'(a)\leqslant\pi'(\tilde{a})$与 $\nu'(a)\leqslant\nu'(\tilde{a})$. 事实上，因为 $a(\lambda)$的友阵 C_a 这时为实矩阵，所以有 $C_a^*=C_a^{\mathrm{T}}$. 应用习题 4.3.1 题 4，B 与 C_a 显然满足下列矩阵方程：

$$(\mathrm{i}C_a)^*B+B(\mathrm{i}C_a)=O. \tag{3}$$

这时，4.2.2 小节引理 8 断言：$\mathrm{In}(-\mathrm{i}C_a^*)\leqslant\mathrm{In}(B)$，因而再应用刚证的结果，$\pi(-\mathrm{i}C_a^*)=\pi(\mathrm{i}C_a)=\nu'(a)\leqslant\pi(B)=\nu(-B)=\nu'(\tilde{a})$，同理可证，$\pi(a)\leqslant\pi'(\tilde{a})$.

最后，假如还有 $\delta'(a)=\delta(\mathrm{i}C_a)=0$，将 4,2,2 小节定理 9 用于方程(3)便知(2)式中等式均成立.　□

上述定理告诉我们，当 $\det B\neq0$ 时，扰动 $\mathrm{i}b(\lambda)$($b(\lambda)$为实多项式)起着将 $a(\lambda)$的实零点“赶出”实轴的作用. 并且，扰动多项式 $\tilde{a}(\lambda)$位于实轴上部与下部的零点个数可用 $\mathrm{Bez}(a,b)$的惯性求得.

例 2　考虑复多项式 $\tilde{a}(\lambda)=\lambda^4+4\mathrm{i}\lambda^3-4\lambda^2-1$. 将 $\tilde{a}(\lambda)$改写为 $\tilde{a}(\lambda)=(\lambda^4-4\lambda^2-1)+\mathrm{i}(4\lambda^3)$，即有 $a(\lambda)=\lambda^4-4\lambda^2-1,b(\lambda)=4\lambda^3$. 应用 $\mathrm{Bez}(a,b)$的 Barnett 因式分解：$\mathrm{Bez}(a,b)=S(a)b(C_a)$，

$$\mathrm{Bez}(a,b)=4S(a)C_a^3=\begin{bmatrix}0&-4&0&1\\-4&0&1&0\\0&1&0&0\\1&0&0&0\end{bmatrix}\begin{bmatrix}0&0&0&4\\4&0&16&0\\0&4&0&16\\16&0&68&0\end{bmatrix}=\begin{bmatrix}0&0&4&0\\0&4&0&0\\4&0&16&0\\0&0&0&4\end{bmatrix}.$$

因此，$\det\mathrm{Bez}(a,b)\neq0$，并且容易求出，$\pi(B)=3$ 与 $\nu(B)=1$. 按定理 1，$\pi'(\tilde{a})=1$ 与 $\nu'(\tilde{a})=3$，即 $\tilde{a}(\lambda)$有一个正虚部的零点与三个负虚部的零点.（注意，$a(\lambda)$有两个实零点与一对共轭复零点）　□

推论 3　设 $a(\lambda),b(\lambda),\tilde{a}(\lambda)$与 B 同定理 1. 若 $B>O(<O)$，则 $\nu'(\tilde{a})=l(\pi'(\tilde{a})=l)$.

现在研究 **Liénard-Chipart 稳定性准则的变型**.

1914 年，Liénard 与 Chipart 在 Routh-Hurwitz 结论（见后面推论 13）的基础上，应用特殊的二次型得到实多项式稳定性准则. 此准则指出，实多项式 $a(\lambda)=\sum_{j=0}^{l}a_j\lambda^j\,(a_l>0)$ 稳定的充分与必要条件为

$$a_0>0,a_2>0,\cdots;\Delta_1>0,\Delta_3>0,\cdots, \tag{4}$$

这里，$\Delta_1,\Delta_3,\Delta_5,\cdots$分别为与 $a(\lambda)$对应的如下 **Hurwitz 矩阵** $\hat{H}$ 的一阶，三阶，五阶，

…前主子式：

$$\hat{H}=\begin{bmatrix} a_{l-1} & a_{l-3} & a_{l-5} & \cdots \\ a_l & a_{l-2} & a_{l-4} & \cdots \\ 0 & a_{l-1} & a_{l-3} & \cdots \\ 0 & a_l & a_{l-2} & \\ 0 & 0 & a_{l-1} & \\ 0 & 0 & a_l & \\ \vdots & \vdots & \vdots & a_0 \end{bmatrix} \in \boldsymbol{M}_l(\mathbb{R}). \tag{5}$$

(注意，$\hat{H}$ 的对角元素为 $a_{l-1},a_{l-2},\cdots,a_0$，且若 $k<0$ 或 $k>l$，则 $a_k=0$)应用定理 1，可以给出 $\mathrm{In}(a)$的表达公式(见定理 4)，因而也得到上述经典的 Liénard-Chipart 准则的一种修正形式(见后面推论 5).

将实多项式 $a(\lambda)=\sum_{j=0}^{l}a_j\lambda^j\ (a_l\neq 0)$ 改写成

$$a(\lambda)=h(\lambda^2)+\lambda g(\lambda^2), \tag{6}$$

式中

$$h(\lambda)=\begin{cases}\sum_{j=0}^{l/2}a_{2j}\lambda^j, l\text{ 为偶数}\\ \sum_{j=0}^{(l-1)/2}a_{2j}\lambda^j, l\text{ 为奇数}\end{cases},\quad g(\lambda)=\begin{cases}\sum_{j=0}^{l/(2-1)}a_{2j+1}\lambda^j, l\text{ 为偶数}\\ \sum_{j=0}^{(l-1)/2}a_{2j+1}\lambda^j, l\text{ 为奇数}\end{cases}. \tag{7}$$

(6)式称为 $a(\lambda)$的 ***L-C* 劈分**(*L-C* splitting).

例如，$a(\lambda)=(\lambda+1)^2(\lambda+2)=2+5\lambda+4\lambda^2+\lambda^3$ 的 *L-C* 劈分中，$h(\lambda)=2+4\lambda$，$g(\lambda)=5+\lambda$，因而

$$a(\lambda)=(2+4\lambda^2)+\lambda(5+\lambda^2)=h(\lambda^2)+\lambda g(\lambda^2). \tag{8}$$

定理 4　设实多项式 $a(\lambda)=\sum_{j=0}^{l}a_j\lambda^j\ (a_l\neq 0)$ 的 *L-C* 劈分如(6)式. 令

$$a_1(\lambda)=\begin{cases}h(-\lambda^2), & l\text{ 为偶数}\\ \lambda g(-\lambda^2), & l\text{ 为奇数}\end{cases},\quad a_2(\lambda)=\begin{cases}\lambda g(-\lambda^2), & l\text{ 为偶数}\\ -h(-\lambda^2), & l\text{ 为奇数}\end{cases}. \tag{9}$$

若 $B=\mathrm{Bez}(a_1,a_2)$非奇异，则有 $\mathrm{In}a=\mathrm{In}B,\delta(a)=\delta(B)=0$.

证明　令 $\tilde{a}(\lambda)\equiv a(\mathrm{i}\lambda)$. 则 $\mathrm{In}'(\tilde{a})=(\nu(a),\pi(a),\delta(a))$. 按(6)式，

$$\tilde{a}(\lambda)=h(-\lambda^2)+\mathrm{i}\lambda g(-\lambda^2).$$

因此，按(9)式，当 l 为偶数时，$\tilde{a}(\lambda)=a_1(\lambda)+\mathrm{i}a_2(\lambda)$；当 l 为奇数时，$-\mathrm{i}\tilde{a}(\lambda)=a_1(\lambda)+\mathrm{i}a_2(\lambda)$. 应用定理 1，并注意到 $\tilde{a}(\lambda)$与 $-\mathrm{i}\tilde{a}(\lambda)$有相同零点以及 $\tilde{a}(\lambda)$没有实零点(因为 $\det B\neq 0$)的事实，我们有

$$\nu(B)=\pi'(\tilde{a})=\nu(a),\qquad \pi(B)=\nu'(\tilde{a})=\pi(a),$$

$$\delta(B)=\delta'(\tilde{a})=\delta(a)=0.$$

□

特别地，若 $B=\mathrm{Bez}(a_1,a_2)<O$，即 $\nu(B)=l$，则前一定理得出 $\nu(a)=l$，因而我们有如下经典的 Liénard-Chipart 稳定性准则的一种修正形式.

推论 5　设 $a(\lambda)$，$a_1(\lambda)$ 与 $a_2(\lambda)$ 同定理 4. 则当且仅当 $B=\mathrm{Bez}(a_1,a_2)<O$ 时，实多项式 $a(\lambda)$ 为稳定的.

证明　按定理 4 我们只要证明：$a(\lambda)$ 稳定蕴涵 $B=\mathrm{Bez}(a_1,a_2)$ 非奇异. 若不然，根据 4.3.1 小节定理 3，$a_1(\lambda)$ 与 $a_2(\lambda)$ 有公共零点 λ_0，因而 $a_1(\bar{\lambda}_0)=a_2(\bar{\lambda}_0)=0$ 与 $\tilde{a}(\lambda_0)=\tilde{a}(\bar{\lambda}_0)=0$，这表明 $a(\lambda)$ 有复零点 $\mu=\mathrm{i}\lambda_0$ 与 $-\bar{\mu}=\mathrm{i}\bar{\lambda}_0$（若 $\lambda_0\neq\bar{\lambda}_0$）或者有一纯虚零点 $\mathrm{i}\lambda_0$（若 $\lambda_0=\bar{\lambda}_0$）. 但这两种情形均与 $a(\lambda)$（关于虚轴）稳定假定矛盾.　□

我们用(8)式中三次实多项式 $a(\lambda)=2+5\lambda+4\lambda^2+\lambda^3$ 来验证一下定理 4. 这时，$l=3$，且 $a_1(\lambda)=\lambda g(-\lambda^2)=5\lambda-\lambda^3$，$a_2(\lambda)=-h(-\lambda^2)=-2+4\lambda^2$，按 Barnett 因式分解，

$$B=\mathrm{Bez}(a_1,a_2)=\begin{bmatrix}5&0&-1\\0&-1&0\\-1&0&0\end{bmatrix}\begin{bmatrix}-2&0&4\\0&18&0\\0&0&18\end{bmatrix}=\begin{bmatrix}-10&0&2\\0&-18&0\\2&0&-4\end{bmatrix}<O.$$

于是，$\mathrm{In}B=\{0,3,0\}$. 按定理 4，$\mathrm{In}(a)=\{0,3,0\}$，即 $a(\lambda)$ 为稳定的. 实际上，$a(\lambda)$ 的零点为 $-1,-1,-2$.

从定理 4 出发，可得到实多项式惯性的另一种表达式，它的好处在于只要计算两个较小阶的 Bézout 矩阵的惯性便能得出实多项式的惯性.

定理 6　设实多项式 $a(\lambda)$ 有 L-C 劈分(6). 若 $B_1=\mathrm{Bez}(h,g)$ 与 $B_2=\mathrm{Bez}(h,g_1)$ 非奇异，这里 $g_1(\lambda)=\lambda g(\lambda)$，则

$$\delta(a)=\delta(B_1)+\delta(B_2)=0,$$
$$\pi(a)=\nu(B_1)+\pi(B_2),$$
$$\nu(a)=\pi(B_1)+\nu(B_2).$$

特别地，当且仅当 $B_1>O$ 与 $B_2<O$ 时，$a(\lambda)$ 为稳定的.

证明　先讨论 l 为偶数即 $l=2k$ 情形. 令 $B=\mathrm{Bez}(a_1,a_2)$，这里，$a_1(\lambda)$ 与 $a_2(\lambda)$ 涵义同(9)式，因而从(7)式看出，

$$a_1(\lambda)=h(-\lambda^2)=\sum_{j=0}^{k+1}(-1)^{j-1}a_{2j-2}\lambda^{2j-2},$$
$$a_2(\lambda)=\lambda g(-\lambda^2)=\sum_{j=1}^{k}(-1)^{j-1}a_{2j-1}\lambda^{2j-1}.$$

应用 B 的 Barnett 因式分解与 4.3.1 小节(8)式，我们有

$$\begin{aligned}B=\mathrm{Bez}(a_1,a_2)&=S(a_1)a_2(C_{a_1})=\sum_{j=1}^{k}(-1)^{j-1}a_{2j-1}S(a_1)C_{a_1}^{2j-1}\\&=\sum_{j=1}^{k}(-1)^{j-1}a_{2j-1}\mathrm{diag}[A_{2j-1}^{(1)},A_{2j-1}^{(2)}],\end{aligned}\tag{10}$$

式中，$A_{2j-1}^{(1)}\in \boldsymbol{M}_{2j-1}(\mathbb{R})$与$A_{2j-1}^{(2)}\in \boldsymbol{M}_{l-2j+1}(\mathbb{R})$有形式

$$A_{2j-1}^{(1)}=\begin{bmatrix} 0 & \cdots & & 0 & -a_0 \\ & & & -a_0 & 0 \\ \vdots & \iddots & & & a_2 \\ & \iddots & & & 0 \\ 0 & & & & \vdots \\ -a_0 & 0 & a_2 & 0 \quad \cdots & (-1)^j a_{2j-2} \end{bmatrix},$$

$$A_{2j-1}^{(2)}=\begin{bmatrix} (-1)^{j+1}a_{2j} & 0 & \cdots & (-1)^k a_l \\ 0 & & \iddots & 0 \\ \vdots & & \iddots & \vdots \\ (-1)^k a_l & 0 & \cdots & 0 \end{bmatrix}.$$

对$j=1,2,\cdots,k$，将(10)式中分块对角阵的行列做相合置换(取$\boldsymbol{Q}=[\boldsymbol{e}_1\ \boldsymbol{e}_3\cdots\ \boldsymbol{e}_{l-1}\ \boldsymbol{e}_2\ \boldsymbol{e}_4\cdots\boldsymbol{e}_l]$)，我们得到

$$\boldsymbol{Q}^{\mathrm{T}}\mathrm{diag}[A_{2j-1}^{(1)},A_{2j-1}^{(2)}]\boldsymbol{Q}=\mathrm{diag}[A_{2j+1}^{[1]},A_{2j+1}^{[2]},A_{2j-1}^{[1]},A_{2j-1}^{[2]}], \tag{11}$$

其中，$A_{2s-1}^{[1]}\in \boldsymbol{M}_{s-1}(\mathbb{R})$与$A_{2s-1}^{[2]}\in \boldsymbol{M}_{k-s+1}(\mathbb{R})(s=j,j+1)$有形式

$$A_{2s-1}^{[1]}=\begin{bmatrix} 0 & \cdots & 0 & -a_0 \\ \vdots & & \iddots & a_2 \\ 0 & \iddots & \iddots & \vdots \\ -a_0 & a_2 & \cdots & (-1)^{s-1}a_{2s-4} \end{bmatrix},$$

$$A_{2s-1}^{[2]}=\begin{bmatrix} (-1)^s a_{2s} & \cdots & & (-1)^k a_l \\ & & & 0 \\ \vdots & \iddots & \iddots & \vdots \\ (-1)^k a_l & 0 & \cdots & 0 \end{bmatrix}.$$

将(11)式代入(10)式，则推出

$$\begin{aligned} Q^{\mathrm{T}}BQ &= \sum_{j=1}^{k}(-1)^{j-1}a_{2j-1}\mathrm{diag}[A_{2j+1}^{[1]},A_{2j+1}^{[2]},A_{2j-1}^{[1]},A_{2j-1}^{[2]}] \\ &\equiv \mathrm{diag}[\widetilde{B}_2,\widetilde{B}_1], \end{aligned} \tag{12}$$

其中$\widetilde{B}_2,\widetilde{B}_1\in \boldsymbol{M}_k(\mathbb{R})$. 但由 4.3.1 小节(8)式看出，上式中$\mathrm{diag}[A_{2j-1}^{[1]},A_{2j-1}^{[2]}]$恰是$S(h(-\lambda))C_{h(-\lambda)}^{j-1}(1\leqslant j\leqslant k)$，同样地，$\mathrm{diag}[A_{2j+1}^{[1]},A_{2j+1}^{[2]}]$恰是$S(h(-\lambda))C_{h(-\lambda)}^{j}(1\leqslant j\leqslant k)$，因而

$$\widetilde{B}_1=S(h(-\lambda))g(-C_{h(-\lambda)})=\mathrm{Bez}(h(-\lambda),g(-\lambda)),$$
$$\widetilde{B}_2=S(h(-\lambda))C_{h(-\lambda)}g(-C_{h(-\lambda)})=\mathrm{Bez}(h(-\lambda),\lambda g(-\lambda)).$$

应用 4.3.1 小节命题 2(1)，可以直接验明

$$\begin{aligned} D\widetilde{B}_1D &= -B_1 = -\mathrm{Bez}(h,g), \\ D\widetilde{B}_2D &= -\mathrm{Bez}(h,-g_1)=\mathrm{Bez}(h,g_1)=B_2, \end{aligned} \tag{13}$$

式中，$D=\mathrm{diag}[1,-1,\cdots,(-1)^{k-1}]\in \boldsymbol{M}_k(\mathbb{R})$. 根据 Sylvester 惯性律与(12)，(13)两式，我们有 $\mathrm{In}B=\mathrm{In}\widetilde{B}_1+\mathrm{In}\widetilde{B}_2=\mathrm{In}(-B_1)+\mathrm{In}B_2$. 再应用定理 4 便得 $\mathrm{In}(a)=\mathrm{In}(-B_1)+\mathrm{In}B_2$ 与 $\delta(a)=0$(因为 B_1 与 B_2 非奇异按(12)与(13)式等价于 $\det B\neq 0$).

对 l 为奇数的情形可类似地讨论(见习题 2). 此外，本定理最后部分为前一部分结果的直接推论. □

例 7　考虑 $a(\lambda)=(\lambda-1)^3(\lambda+2)$ 的惯性 $\mathrm{In}(a)$. 这时，$h(\lambda)=-2-3\lambda+\lambda^2$，$g(\lambda)=5-\lambda$，并且

$$B_1=\mathrm{Bez}(h,g)=\begin{bmatrix}-17 & 5\\ 5 & -1\end{bmatrix} \text{与}\ B_2=\mathrm{Bez}(h,g_1)=\begin{bmatrix}10 & -2\\ -2 & 2\end{bmatrix}.$$

显然，$\mathrm{In}B_1=\{1,1,0\}$ 与 $\mathrm{In}B_2=\{2,2,0\}$，根据定理 6，$\mathrm{In}(a)=\mathrm{In}(-B_1)+\mathrm{In}B_2=\{3,1,0\}$. 实际上，$a(\lambda)$ 的零点为 $1,1,1,-2$. □

应用与实多项式 $a(\lambda)$ 有关的 l 阶 Hankel 矩阵与 Bézout 矩阵之间的相合关系，我们还可以推出定理 4 与 6 的等价形式，即 **Markov 准则**. 为此，先引出 **Markov 参数的概念**.

设 $b(\lambda)=\sum\limits_{j=0}^{l}b_j\lambda^j$ 与 $c(\lambda)=\sum\limits_{j=0}^{l}c_j\lambda^j\ (c_l\neq 0)$ 均为实多项式. 考虑 λ 的有理函数 $R(\lambda)=b(\lambda)/c(\lambda)$，显然有 $\lim\limits_{\lambda\to\infty}|R(\lambda)|<\infty$. 因此，对充分大的 $|\lambda|$，$R(\lambda)$ 可展成 Laurent 级数：

$$R(\lambda)=\sum_{j=0}^{\infty}h_j\lambda^{-j}. \tag{14}$$

上式中 $h_0,h_1,h_2,\cdots$ 称为 $R(\lambda)$ 的 **Markov 参数**. 实数列 $\{h_j\}_{j=1}^{\infty}$ 确定出一个无穷阶的 Hankel 矩阵 $[h_{i+j-1}]_{i,j=1}^{\infty}$. 今后，用 H 表示这个无穷阶矩阵的第 l 个前主子阵，即 $H=[h_{i+j-1}]_{i,j=1}^{l}\in \boldsymbol{M}_l(\mathbb{R})$. 显然，$H=H^{\mathrm{T}}$.

有趣的是 $B=\mathrm{Bez}(c,b)$ 与 $R(\lambda)=b(\lambda)/c(\lambda)$ 的 Markov 参数组成的 Hankel 矩阵 $H\in \boldsymbol{M}_l(\mathbb{R})$ 是相合的，因而 $\mathrm{In}B=\mathrm{In}H$. 确切地我们有

命题 8　设 $b(\lambda)=\sum\limits_{j=0}^{l}b_j\lambda^j$ 与 $c(\lambda)=\sum\limits_{j=0}^{l}c_j\lambda^j\ (c_l\neq 0)$ 为实多项式. 则有

$$\mathrm{Bez}(c,b)=S(c)HS(c), \tag{15}$$

式中，$H=[h_{i+j-1}]_{i,j=1}^{l}\in \boldsymbol{M}_l(\mathbb{R})$ 是 $R(\lambda)=b(\lambda)/c(\lambda)$ 的 Markov 参数组成的 l 阶 Hankel 矩阵，$S(c)$ 为多项式 $c(\lambda)$ 的对称化子.

证明　按定义有

$$b(\lambda)=(h_0+h_1\lambda^{-1}+h_2\lambda^{-2}+\cdots)c(\lambda). \tag{16}$$

比较上式两端的对应项系数便得

$$b_l=h_0c_l,$$

$$\begin{aligned} b_{l-1} &= h_1 c_l + h_0 c_{l-1}, \\ b_{l-2} &= h_2 c_l + h_1 c_{l-1} + h_0 c_{l-2}, \\ &\vdots \\ b_0 &= h_l c_l + h_{l-1} c_{l-1} + \cdots + h_1 c_1 + h_0 c_0, \end{aligned} \tag{17}$$

以及

$$0 = h_{j+l} c_l + h_{j+l-1} c_{l-1} + \cdots + h_{j+1} c_1 + h_j c_0, \qquad j = 1, 2, \cdots. \tag{18}$$

今令 $\boldsymbol{\beta}^{\mathrm{T}} = (\tilde{b}_0\ \tilde{b}_1 \cdots \tilde{b}_{l-1})$，$\tilde{b}_j = b_j - b_l c_j / c_l$，$j = 0, 1, \cdots, l-1$；$\boldsymbol{h}_1^{\mathrm{T}} = (h_1\ h_2 \cdots h_l)$，由(17)式可得

$$\boldsymbol{\beta}^{\mathrm{T}} = \boldsymbol{h}_1^{\mathrm{T}} S(c). \tag{19}$$

应用(18)式与 4.3.1 小节(8)式，对 $1 \leqslant j \leqslant l-1$，我们有

$$\begin{aligned} \boldsymbol{\beta}^{\mathrm{T}} C_c^j &= \boldsymbol{h}_1^{\mathrm{T}} S(c) C_c^j = \boldsymbol{h}_1^{\mathrm{T}} \operatorname{diag}\left[-\begin{bmatrix} 0 & & c_0 \\ & \cdot^{\cdot^{\cdot}} & \vdots \\ c_0 & \cdots & c_{j-1} \end{bmatrix}, \begin{bmatrix} c_{j+1} & \cdots & c_l \\ \vdots & \cdot^{\cdot^{\cdot}} & \\ c_l & & 0 \end{bmatrix} \right] \\ &= (-c_0 h_j \cdots - c_0 h_1 - \cdots - c_{j-1} h_j \quad c_{j+1} h_{j+1} + \cdots + c_l h_l \cdots c_l h_{j+1}) \\ &= \boldsymbol{h}_{j+1}^{\mathrm{T}} S(c), \end{aligned} \tag{20}$$

式中，$\boldsymbol{h}_{j+1}^{\mathrm{T}} = (h_{j+1}\ h_{j+2} \cdots h_{j+l})$ 为 $H \in \boldsymbol{M}_l(\mathbb{R})$ 的第 $j+1$ 个行向量，此时，(19)，(20)式蕴涵

$$\begin{bmatrix} \boldsymbol{\beta}^{\mathrm{T}} \\ \boldsymbol{\beta}^{\mathrm{T}} C_c \\ \vdots \\ \boldsymbol{\beta}^{\mathrm{T}} C_c^{l-1} \end{bmatrix} = HS(c).$$

于是，(15)式可从 $\mathrm{Bez}(c,b)$ 的 Barnett 因式分解与上式直接推得. □

对于实多项式 $a(\lambda) = \sum_{j=0}^{l} a_j \lambda^j \ (a_l \neq 0)$，若取命题 8 中 $b(\lambda)$ 与 $c(\lambda)$ 分别为定理 4 中的 $a_2(\lambda)$ 与 $a_1(\lambda)$，那么便得 $B = \mathrm{Bez}(a_1, a_2) = S(a_1) HS(a_1)$，这里 H 为对应于 $a_2(\lambda)/a_1(\lambda)$ Markov 参数的 l 阶 Hankel 矩阵. 应用定理 4，推论 5 与 Sylvester 惯性律，我们便得到 Markov 稳定性准则的如下修正形式.

定理 9　设 $a(\lambda)$，$a_1(\lambda)$ 与 $a_2(\lambda)$ 同定理 4，H 为由 $a_2(\lambda)/a_1(\lambda)$ Markov 参数组成的 l 阶 Hankel 矩阵. 若 $\det H \neq 0$，则 $\mathrm{In}(a) = \mathrm{In} H$，$\delta(a) = \delta(H) = 0$，特别地，当且仅当 $H < O$ 时，$a(\lambda)$ 为稳定的.

我们用定理 9 再次来验证(8)式中实多项式 $a(\lambda) = (\lambda+1)^2(\lambda+2)$ 的稳定性. 事实上，

$$R(\lambda) = \frac{a_2(\lambda)}{a_1(\lambda)} = \frac{-2 + 4\lambda^2}{5\lambda - \lambda^3} = -4\lambda^{-1} - 18\lambda^{-3} - 90\lambda^{-5} + \cdots,$$

因而对应的 Hankel 矩阵 $H \in \boldsymbol{M}_3(\mathbb{R})$ 为

$$H=\begin{bmatrix}-4 & 0 & -18\\ 0 & -18 & 0\\ -18 & 0 & -90\end{bmatrix}.$$

容易验证 $H<O$,按定理 9,$a(\lambda)$为稳定的.

同理,若将命题 8 与 Sylvester 惯性律用于定理 6 中的 Bézout 矩阵 B_1 与 B_2,则我们便得如下的 **Markov 稳定性准则**.

定理 10(Markov 稳定性准则) 设 $a(\lambda),h(\lambda),g(\lambda),B_1$ 与 B_2 同定理 6,又设 Markov 参数 $h_0,h_1,h_2,\cdots$由下式确定:对充分大的$|\lambda|$,

$$\sum_{j=0}^{\infty}h_j\lambda^{-j}=\begin{cases}g(\lambda)/h(\lambda), & l\text{ 为偶数},\\ h(\lambda)/(\lambda g(\lambda)), & l\text{ 为奇数}.\end{cases}\tag{21}$$

若如下两个 Hankel 矩阵 H_1 与 H_2 非奇异:

$$H_1=[h_{i+j-1}]_{i,j=1}^{k},H_2=[h_{i+j}]_{i,j=1}^{m},\tag{22}$$

这里,当 l 偶数时,$k=m=l/2$;当 l 奇数时,$k=(l+1)/2$ 与 $m=(l-1)/2$,则有

$$\begin{aligned}\pi(a)&=\nu(H_1)+\pi(H_2),\\ \nu(a)&=\pi(H_1)+\nu(H_2),\\ \delta(a)&=\delta(H_1)+\delta(H_2)=0.\end{aligned}\tag{23}$$

特别地,当且仅当 $H_1>O$ 与 $H_2<O$ 时,$a(\lambda)$为稳定的.

证明 当 l 为偶数时,(21)式中 $h_0=0$. 这时由于 $h(\lambda)$为 $l/2$ 次实多项式,故(15)式蕴涵 $B_1=\mathrm{Bez}(h,g)$与(22)式中的 H_1 相合. 另一方面,由于对充分大$|\lambda|$,$\lambda g(\lambda)/h(\lambda)=\sum_{j=0}^{\infty}h_{j+1}\lambda^{-j}$ 成立,故(15)式蕴涵 $B_2=\mathrm{Bez}(h,g_1)(g_1(\lambda)=\lambda g(\lambda))$与(22)式中的 H_2 相合. 因此 $\mathrm{In}B_1=\mathrm{In}H_1$ 与 $\mathrm{In}B_2=\mathrm{In}H_2$. 再引用定理 6 便得(23)式.

当 l 为奇数时,$h(\lambda)$的次数不超过 $g(\lambda)$的次数,且 $g(\lambda)$为$(l-1)/2$ 次多项式. 应用(21)式($h_0=0$),我们有

$$h(\lambda)/g(\lambda)=\sum_{j=0}^{\infty}h_{j+1}\lambda^{-j}\text{ 与 }h(\lambda)/(\lambda g(\lambda))=\sum_{j=1}^{\infty}h_j\lambda^{-j}=\sum_{j=0}^{\infty}h_{j+1}\lambda^{-j+1},$$

因而,按命题 8,H_2 与$-B_1=\mathrm{Bez}(g,h)$相合,H_1 与$-B_2=\mathrm{Bez}(g_1,h)$相合($g_1(\lambda)=\lambda g(\lambda)$). 因此,$\mathrm{In}H_2=\mathrm{In}(-B_1)$与 $\mathrm{In}H_1=\mathrm{In}(-B_2)$. 再引用定理 6 便得(23)式. □

将定理 10 用于(8)式中多项式 $a(\lambda)=(\lambda+1)^2(\lambda+2)$. 这时,

$$\frac{h(\lambda)}{\lambda g(\lambda)}=\frac{2+4\lambda}{\lambda(5+\lambda)}=4\lambda^{-1}-18\lambda^{-2}+90\lambda^{-3}+\cdots,$$

于是,

$$H_1=\begin{bmatrix}h_1 & h_2\\ h_2 & h_3\end{bmatrix}=\begin{bmatrix}4 & -18\\ -18 & 90\end{bmatrix}\quad\text{与}\quad H_2=[h_2]=[-18].$$

显然,$H_1>O$ 与 $H_2<O$. 按定理 10,$a(\lambda)$为稳定的.

最后,我们讨论 **Routh-Hurwitz 稳定性定理.**

还考虑实多项式 $a(\lambda)=\sum_{j=0}^{l}a_l\lambda^j\,(a_l\neq 0)$,以及与 $a(\lambda)$ 对应的 Hurwitz 矩阵 $\hat{H}\in \boldsymbol{M}_l(\mathbb{R})$(见(5) 式). 一般地,$\hat{H}$ 不是对称的,但是下面将要证明:$a(\lambda)$ 的 Hurwitz 矩阵 $\hat{H}$ 的前主子式与 $g(\lambda)/h(\lambda)$ 的 Markov 参数组成的 l 阶 Hankel 矩阵 $H(H=H^{\mathrm{T}})$的前主子式之间有着明显的关系(详见定理12的证明),这里,$a(\lambda)=h(\lambda^2)+\lambda g(\lambda^2)$ 为 $a(\lambda)$ 的 L-C 劈分. 为此先研究一个更一般的命题.

命题 11　设 $b(\lambda),c(\lambda)$与 $H=[h_{i+j-1}]_{i,j=1}^{l}$同命题 8. 若约定 $b_j=c_j=0,\forall j<0$ 或 $j>l$,则对 $1\leqslant s\leqslant l$,

$$c_l^{2s}\det[h_{i+j-1}]_{i,j=1}^{s}=\det\begin{bmatrix} c_l & c_{l-1} & \cdots & & c_{l-2s+1}\\ b_l & b_{l-1} & \cdots & & b_{l-2s+1}\\ 0 & c_l & c_{l-1} & \cdots & c_{l-2s+2}\\ 0 & b_l & b_{l-1} & \cdots & b_{l-2s+2}\\ \vdots & \vdots & \vdots & & \vdots\\ 0 & \cdots & & c_l & \cdots c_{l-s}\\ 0 & \cdots & & b_l & \cdots \end{bmatrix}. \tag{24}$$

证明　对 $0\leqslant k\leqslant 2s-1$,引入两个 $1\times 2s$ 矩阵

$$\boldsymbol{h}_k^{\mathrm{T}}=[0\ \ \cdots\ \ 0\ \ h_0\ \ h_1\ \ \cdots\ \ h_k],$$
$$\boldsymbol{b}_k^{\mathrm{T}}=[0\ \ \cdots\ \ 0\ \ b_l\ \ b_{l-1}\ \ \cdots\ \ h_{l-k}].$$

应用(17)式,对每个 k,我们有

$$\boldsymbol{h}_k^{\mathrm{T}}\begin{bmatrix} c_l & c_{l-1} & \cdots & c_{l-2s+1}\\ 0 & c_l & \ddots & \vdots\\ \vdots & \ddots & \ddots & c_{l-1}\\ 0 & \cdots & 0 & c_l \end{bmatrix}=\boldsymbol{b}_k^{\mathrm{T}}. \tag{25}$$

令 $\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_{2s}$为$\mathbb{C}^{2s}$中的单位坐标向量,且 $\boldsymbol{c}_k^{\mathrm{T}}=[0\ \ \cdots\ \ 0\ \ c_l\ \ c_{l-1}\ \ \cdots\ \ c_{l-k}]$ 为 $1\times 2s$ 矩阵,$0\leqslant k\leqslant 2s-1$. 约定 $c_j=b_j=0,\forall j>l$ 或$j<0$. 由(25)式推出,

$$\begin{bmatrix} \boldsymbol{e}_1^{\mathrm{T}}\\ \boldsymbol{h}_{2s-1}^{\mathrm{T}}\\ \boldsymbol{e}_2^{\mathrm{T}}\\ \boldsymbol{h}_{2s-2}^{\mathrm{T}}\\ \vdots\\ \boldsymbol{e}_s^{\mathrm{T}}\\ \boldsymbol{h}_s^{\mathrm{T}} \end{bmatrix}\begin{bmatrix} c_l & c_{l-1} & \cdots & c_{l-2s+1}\\ 0 & c_l & \ddots & \vdots\\ \vdots & \ddots & \ddots & c_{l-1}\\ 0 & \cdots & 0 & c_l \end{bmatrix}=\begin{bmatrix} \boldsymbol{c}_{2s-1}^{\mathrm{T}}\\ \boldsymbol{b}_{2s-1}^{\mathrm{T}}\\ \boldsymbol{c}_{2s-2}^{\mathrm{T}}\\ \boldsymbol{b}_{2s-2}^{\mathrm{T}}\\ \vdots\\ \boldsymbol{c}_s^{\mathrm{T}}\\ \boldsymbol{b}_s^{\mathrm{T}} \end{bmatrix}. \tag{26}$$

上式右端矩阵的行列式即为(24)式的右端项，而左端的两个 $2s$ 阶矩阵乘积的行列式等于 $c_l^{2s}\det[h_{i+j-1}]_{i,j=1}^{s}$，因而(24)式成立. □

我们指出，(24)式的左端与右端行列式的阶数分别为 s 与 $2s$.

有了命题 11，应用定理 10 与 Jacobi 法则(见第一章习题 1.2.1 题 7)，我们可以导出如下著名的 **Routh-Hurwitz 稳定性定理**.

定理 12 设 $a(\lambda)=\sum_{j=0}^{l}a_j\lambda^j\ (a_l\neq 0)$ 为实多项式，若 $a(\lambda)$ 的 Hurwitz 矩阵 $\hat{H}$ 的头 l 个前主子式满足.

$$\Delta_1\neq 0,\Delta_2\neq 0,\cdots,\Delta_l\neq 0, \tag{27}$$

则 $\delta(a)=0$，且

$$\begin{aligned}\pi(a)&=V\left(a_l,\Delta_1,\frac{\Delta_2}{\Delta_1},\cdots,\frac{\Delta_l}{\Delta_{l-1}}\right),\\ \nu(a)&=l-\pi(a)=P\left(a_l,\Delta_1,\frac{\Delta_2}{\Delta_1},\cdots,\frac{\Delta_l}{\Delta_{l-1}}\right),\end{aligned} \tag{28}$$

这里，V 与 P 分别表示对应数列符号变化的次数(即变号数)与符号不变的次数(即连号数).

证明 设(6)式为 $a(\lambda)$的 L-C 劈分. 考虑 l 为偶数情形：$l=2k$. 选取

$$b(\lambda)=\lambda g(\lambda)=\sum_{j=0}^{k-1}a_{2j+1}\lambda^{j+1}\quad 与\quad c(\lambda)=h(\lambda)=\sum_{j=0}^{k}a_{2j}\lambda^j.$$

若 $h_0=0,h_1,h_2,\cdots$表示 $g(\lambda)/h(\lambda)$的 Markov 参数，则 $b(\lambda)/c(\lambda)$的 Markov 参数为 $h_1,h_2,\cdots$，且(24)式右端对应的矩阵为

$$\begin{bmatrix} a_l & a_{l-2} & \cdots & \cdots & a_{l-4s+2}\\ a_{l-1} & a_{l-3} & \cdots & \cdots & a_{l-4s+1}\\ 0 & a_l & a_{l-2} & \cdots & a_{l-4s+4}\\ 0 & a_{l-1} & a_{l-3} & \cdots & a_{l-4s+3}\\ 0 & 0 & a_l & \cdots & \\ 0 & 0 & a_{l-1} & \cdots & \\ \vdots & \vdots & \vdots & & \end{bmatrix}\in \boldsymbol{M}_{2s}(\mathbb{R}),\qquad 1\leqslant s\leqslant k.$$

这里约定 $a_j=0,\forall j<0$. 交换上述矩阵的 s 行之后，它变为(5)式中的矩阵 $\hat{H}$ 的 $2s$ 阶前主子阵. 由(24)式推出，

$$a_l^{2s}\delta_s^{(2)}=(-1)^s\Delta_{2s},\qquad s=1,2,\cdots,k, \tag{29}$$

式中，$\delta_s^{(2)}$ 为(22)式中 H_2 的 s 阶前主子式 $\det[h_{i+j}]_{i,j=1}^{s}$(因为 $b(\lambda)/c(\lambda)$的 Markov 参数为 $h_1,h_2,h_3,\cdots$)，而 Δ_{2s}表示 Hurwitz 矩阵 $\hat{H}$ 的 $2s$ 阶前主子式.

如在命题 11 中取 $b(\lambda)=g(\lambda)$与 $c(\lambda)=h(\lambda)$，则 $b(\lambda)/c(\lambda)$的 Markov 参数恰为 $h_0,h_1,h_2,\cdots$. 应用(24)式便得

$$a_l^{2s}\delta_s^{(1)} = \det\begin{bmatrix} a_l & a_{l-2} & \cdots & \cdots & a_{l-4s+2} \\ 0 & a_{l-1} & a_{l-3} & \cdots & a_{l-4s+3} \\ 0 & a_l & a_{l-2} & \cdots & a_{l-4s+4} \\ 0 & 0 & a_{l-1} & \cdots & a_{l-4s+5} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & & \end{bmatrix} = a_l\Delta_{2s-1}, \qquad 1 \leqslant s \leqslant k, \quad (30)$$

式中，$\delta_s^{(1)}$ 为(22)式中 H_1 的 s 阶前主子式，而 Δ_{2s-1} 表示 Hurwitz 矩阵 $\hat{H}$ 的 $2s-1$ 阶前主子式.

应用定理 10 与 Jacobi 法则(见第一章习题 1.2.1 题 7)，若 $\Delta_1,\Delta_2,\cdots,\Delta_l$ 均不为零，由(29)与(30)式我们有 $\delta(a)=0$，并且，

$$\pi(a) = \nu(H_1) + \pi(H_2) = V(1,\delta_1^{(1)},\cdots,\delta_k^{(1)}) + P(1,\delta_1^{(2)},\cdots,\delta_k^{(2)}) \quad . (31)$$

当 $a_l>0$ 时，上式可改写为

$$\begin{aligned}\pi(a) &= V(1,\Delta_1,\Delta_3,\cdots,\Delta_{l-1}) + P(1,-\Delta_2,\Delta_4,\cdots,(-1)^k\Delta_l) \\ &= V(1,\Delta_1,\Delta_3,\cdots,\Delta_{l-1}) + V(1,\Delta_2,\Delta_4,\cdots,\Delta_l) \\ &= V\left(a_l,\Delta_1,\frac{\Delta_2}{\Delta_1},\cdots,\frac{\Delta_l}{\Delta_{l-1}}\right).\end{aligned}$$

同时，

$$\nu(a) = l - \pi(a) = P\left(a_l,\Delta_1,\frac{\Delta_2}{\Delta_1},\cdots,\frac{\Delta_l}{\Delta_{l-1}}\right).$$

对于 l 为偶数而 $a_l<0$ 情形以及 l 为奇数的情形可按上述类似方法证明(28)式也成立(见习题 3). □

推论 13　设 $a(\lambda)=\sum_{j=0}^{l} a_j\lambda^j\ (a_l>0)$ 为实多项式，则 $a(\lambda)$ 稳定的充分与必要条件为

$$\Delta_1 > 0, \Delta_2 > 0, \cdots, \Delta_l > 0. \quad (32)$$

例 14　应用定理 12 研究例 7 中实多项式 $a(\lambda)=(\lambda-1)^3(\lambda+2)$ 的惯性. 这时，对应的 Hurwitz 矩阵 $\hat{H}$ 为

$$\hat{H} = \begin{bmatrix} -1 & 5 & 0 & 0 \\ 1 & -3 & -2 & 0 \\ 0 & -1 & 5 & 0 \\ 0 & 1 & -3 & -2 \end{bmatrix}.$$

因此，$\Delta_1=-1,\Delta_2=-2,\Delta_3=-8,\Delta_4=16$. 应用定理 12，可求得

$$\pi(a) = V(1,-1,2,4,-2) = 3, \qquad \nu(a) = 1, \qquad \delta(a) = 0. \qquad \square$$

习题 4.3.3

1. 设实多项式 $a(\lambda)=\sum_{j=0}^{l}a_j\lambda^j\ (a_l\neq 0)$，试证：多项式 $\tilde{a}(\lambda)=a(\lambda)+ib_0\ (b_0>0)$ 没有实零点，且

$$\nu'(a)-\pi'(a)=\begin{cases}0, & \text{如 } l \text{ 为偶数,}\\ \operatorname{sgn}a_l, & \text{如 } l \text{ 为奇数.}\end{cases}$$

2. 对 l 为奇数情形证明定理 6.

3. 对 l 为偶数而 $a_l<0$ 的情形与 l 为奇数情形分别地证明定理 12.

4. 设 $a(\lambda)=(\lambda-1)^2(\lambda^2+1)=\lambda^4-2\lambda^3+2\lambda^2-2\lambda+1$. 试分别地用

(1)定理 4，

(2)定理 6，

(3)定理 9，

(4)定理 12，

求 $\mathrm{In}(a)$.

5. 设 $C=(c_{j+k})\in \boldsymbol{M}_n(\mathbb{C})$ 为奇异的 Hankel 矩阵，$r(<n)$ 为非负整数. 试证下列两个命题等价：

(1)矩阵 C 的头 r 列线性无关，而头 $r+1$ 列线性相关.

(2)矩阵 C 的非奇异前主子阵的最大阶数为 r.

6. 设 $l\geqslant 1$，$a(\lambda)$ 与 $b(\lambda)$ 的二次数不超过 l 且至少有一个为 l 次多项式，它们没有公共零点. 试证：存在数量 β 使得 $a(\lambda)+\beta b(\lambda)$ 有 l 个单垂零点.

7. 设 B 为非奇异矩阵. 试证下列两命题等价：

(1)B 为 Bézout 矩阵.

(2)B^{-1} 为 Hankel 矩阵.

参考文献

[1] Lancaster P, Tismenetsky M. The Theory of Matrices, with Applications. 2nd Ed. New York: Academic Press, 1985

[2] Gantmacher F R. Applications of the Theory of Matrices. New York: Interscience, 1959

[3] Bellman R. Introduction to Matrix Analysis. 2nd ed, New York: McGraw-Hill, 1970; Philadelphia: SIAM Press, 1995

[4] Barnett S, Storey C. Matrix Methods in Stability Theory. London: Nelson, 1970

[5] Fiedler M. Special Matrices and their Applications in Numerical Mathematics. Boston: Martinus Nijhoff Publ., 1986

[6] Carlson D, Hill R D. Controllability and inertia theory for function of matrix. J Math. Anal. Appl., 1977(59), 260～266

[7] Cain B E. Inertia theory. Linear Algebra Appl., 1980(30). 211～240; Corrections in Linear Algebra Appl., 1982(42), 285～286

[8] Carlson D, Datta B N. On the effective computation of the inertia of a nonhermitian matrix. Numer. Math., 1979(33), 315～332

[9] Graham A. Kronecker Products and Matrix Calculus with Applications. Horwood, Chichester, U, K. 1981

[10] Taussky O. A generalization of a theorem of Lyapunov. SIAM J. , 1961(9), 640～643

[11] Daleckii L, Krein M G . Stability of Solutions of differential equations in Banach Space. Amer. Math. Soc. , Providence, 1974

[12] Ostrowski A, Schneider H. Some theorems on the inertia of general matrices. J. Math . Anal. Appl. , 1962(4), 72～84

[13] Lancaster P, Tismenetsky M. Inertia characteristics of self-adjoint matrix polynomials. Linear Algebra Appl, 1983(52/53), 479～496

[14] Pták V. Lyapunov, Bézout , and Hankel. Linear Algebra. Appl. , 1984(58), 363～390.

[15] Wimmer H K. Generalizations of theorms of Lyapunov and Stein. Linear Algebra. Appl, 1975(10), 139～146

[16] Datta B N. On the Routh-Hurwitz-Fujiwara and the Schur-Cohn-Fujiwara theorems for the root-separation problem. Linear Algebra Appl. , 1978(22), 235～246

[17] Barnett S. Location of zeros of a complex polynomial. Linear Algebra. Appl. , 1970(4), 71～76

[18] Wimmer H K. Inertia theorems for matrices, controllability and linear vibrations. Linear Algebra. Appl. , 1974(8), 337～344

[19] Carlson D, Schneider H. Inertia theorems for matrices: the semidefinite case. J Math. Anal. Appl. , 1963(6), 430～446

[20] Barnett S. Polynomials and Linear Control Systms. New York: Marcel Dekker, 1983

[21] Lerer L, Tismenetsky M. The Bezoutian and the eigenvalue-separation problem for matrix polynomials. Integral Eq. and Operator Theory, 1982(5), 386～445

[22] Heinig G, Rost K. Algebraic Methods for Toeplitz-Like Matrices and Operators. Berlin: Akademic Verlag. 1984

[23] Krein M G, Naimark M A. The method of symmetric and Hermitian forms in the theory of the separation of the roots of algebraic equations. Linear and Multilinear Algebra, 1981(10), 265～308

[24] Chen C T. A generalization of the inertia theorem. SIAM J. Appl. Math. , 1973(15), 158～161

[25] Barnett S. Matrices: Methods and Applications. Oxford: Clarendon Press, 1990

第五章　矩阵的广义逆

若 $A\in \boldsymbol{M}_n(\mathbb{C})$ 非奇异,则存在唯一的 $X\in \boldsymbol{M}_n(\mathbb{C})$ 满足 $AX=XA=I_n$,X 称为 A 的逆,记为 A^{-1}. 这时,线性方程组 $A\boldsymbol{x}=\boldsymbol{b}$ 对任意给定的 $\boldsymbol{b}\in\mathbb{C}^n$ 有唯一解 $\boldsymbol{x}=A^{-1}\boldsymbol{b}$. 若 A 奇异或为长方阵,则上述的 X 不存在. 但是,这时方程组 $A\boldsymbol{x}=\boldsymbol{b}$ 对任意 $\boldsymbol{b}\in \mathrm{Im}A$ 仍然有解,即使对 $\boldsymbol{b}\notin \mathrm{Im}A$,$A\boldsymbol{x}=\boldsymbol{b}$ 为矛盾方程组的情形,我们还可以考虑所谓最小二乘解 $\boldsymbol{x}$,即寻求 $\boldsymbol{x}$ 使得 $A\boldsymbol{x}$ 与 $\boldsymbol{b}$ 在欧氏向量范数 $\|\cdot\|_2$ 意义下最接近:$\|A\boldsymbol{x}-\boldsymbol{b}\|_2=\min$. 因此,自然希望推广原先矩阵逆的概念,使得方程组 $A\boldsymbol{x}=\boldsymbol{b}$ 在前述几种意义下的解都能用推广后的矩阵逆来表示,并且,这些矩阵广义逆应保持原先矩阵逆的一部分性质,特别当 A 非奇异时,这些广义逆必定含有 A^{-1} 或退化为 A^{-1}.

1920 年,E. H. Moore 对 $m\times n$ 矩阵首次给出广义逆的概念. 若 $A\in \boldsymbol{M}_{m,n}(\mathbb{C})$,则 Moore 意义下的广义逆等价于满足

$$AX = P_{\mathrm{Im}A} \text{ 与 } XA = P_{\mathrm{Im}X}$$

的唯一矩阵 $X\in \boldsymbol{M}_{n,m}(\mathbb{C})$,这里,$P_G$ 表示在子空间 G 上的正交投影. 1955 年,对任意给定的 $A\in \boldsymbol{M}_{m,n}(\mathbb{C})$,R. Penrose 证明前述 Moore 意义下的广义逆正是下列所谓 **Penrose 方程**的唯一解 X:

$$AXA = A, \tag{I}$$

$$XAX = X, \tag{II}$$

$$AX = (AX)^*, \tag{III}$$

$$XA = (XA)^*. \tag{IV}$$

这个唯一解 X 通常称为矩阵 A 的 Moore-Penrose 逆. 一年后,他又证明:对于这个唯一的 X,$X\boldsymbol{b}$ 便是前述线性最小二乘问题 $\|A\boldsymbol{x}-\boldsymbol{b}\|_2=\min$ 的具有最小欧氏范数的解(见 5.1.3 小节定理 3). 此后,半个世纪以来,矩阵广义逆的研究获得迅速的发展,并出现了多种形式与背景的矩阵广义逆.

本章旨在介绍两类矩阵广义逆的基本知识,其中一类为基于 Penrose 方程(Ⅰ)～(Ⅳ)的一般矩阵的 λ-逆,另一类为方阵的谱广义逆.

5.1　基于 Penrose 方程的 λ-逆

5.1.1　基本概念与{1}-逆

对于一般矩阵 $A\in \boldsymbol{M}_{m,n}(\mathbb{C})$,应用 Penrose 方程(I)～(IV)的全部或部分可以

定义 A 的一类广义逆. 确切地有

定义 1　设 $A\in \boldsymbol{M}_{m,n}(\mathbb{C})$，$\lambda$ 为集合$\{1,2,3,4\}$的某个不空子集. 我们称矩阵 $X\in \boldsymbol{M}_{n,m}(\mathbb{C})$是 A 的一个**λ-逆**，假如它满足第 i 个 Penrose 方程，$\forall i\in\lambda$. 今后，用 $A\{\lambda\}$表示 A 的所有λ-逆的集合，用 $A^{[\lambda]}$表示 $A\{\lambda\}$中的任一元素.

这样一来，$A^{[1,4]}$表示满足 Penrose 方程(I)与(IV)的 A 的$\{1,4\}$-逆，等等. 显然当$\lambda\subset\mu\subset\{1,2,3,4\}$且$\lambda$不空时，有 $A\{\mu\}\subset A\{\lambda\}$. 譬如，$X\in A\{1,4\}$蕴涵 $X\in A\{1\}$，等等. 若 $1\in\lambda$，则当且仅当 A 为 $m\times n$ 零矩阵时，$n\times m$ 零矩阵为 A 的一个λ-逆.

应用非零且秩为 r 的矩阵 $A\in \boldsymbol{M}_{m,n}(\mathbb{C})$的奇异值分解(见第一章 1.2.2 小节定理 6)

$$A=U\begin{bmatrix}\Sigma & O\\ O & O\end{bmatrix}_{m\times n}V^*,\tag{1}$$

这里，$U\in\mathscr{U}_m$，$V\in\mathscr{U}_n$，$\Sigma=\mathrm{diag}[\sigma_1,\cdots,\sigma_r]$，$\sigma_1,\cdots,\sigma_r$ 为 A 的所有正奇异值，可以直接验证：

$$V\begin{bmatrix}\Sigma^{-1} & O\\ O & O\end{bmatrix}_{n\times m}U^*\in A\{1,2,3,4\}.\tag{2}$$

因此，定义 1 中 A 的各种λ-逆(共 15 种)均存在.

下列推证表明(2)式中矩阵为 $A\{1,2,3,4\}$的唯一元素：若 $X,Y\in A\{1,2,3,4\}$，按 Penrose 方程则有

$$\begin{aligned}X&=X(AX)^*=XX^*A^*=XX^*(AYA)^*=X(AX)^*(AY)^*\\&=X(AX)(AY)=XAY=(XA)^*(YA)^*Y=A^*X^*A^*Y^*Y\\&=A^*Y^*Y=(YA)^*Y=YAY=Y.\end{aligned}$$

通常，将 $A\{1,2,3,4\}$中唯一元素称为 A 的 **Moore-Penrose 逆**，记为 $A^\dagger$. 由(2)式我们有：若(1)式为 $A\in\boldsymbol{M}_{m,n}(\mathbb{C})$的奇异值分解，则

$$A^\dagger=V\begin{bmatrix}\Sigma^{-1} & O\\ O & O\end{bmatrix}U^*\in\boldsymbol{M}_{n,m}(\mathbb{C}).\tag{3}$$

定理 2　设 $A\in\boldsymbol{M}_{m,n}(\mathbb{C})$. 则 A 的各种λ-逆均存在，且 A 只有唯一的$\{1,2,3,4\}$-逆，它由(3)式确定.

例 3　矩阵 $A\in\boldsymbol{M}_{3,2}(\mathbb{C})$：

$$A=\begin{bmatrix}1 & 0\\ 0 & 1\\ 1 & 0\end{bmatrix}$$

的奇异值分解为

$$A=\begin{bmatrix}1/\sqrt{2} & 0 & 1/\sqrt{2}\\ 0 & 1 & 0\\ 1/\sqrt{2} & 0 & -1/\sqrt{2}\end{bmatrix}\begin{bmatrix}\sqrt{2} & 0\\ 0 & 1\\ 0 & 0\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix},$$

按(3)式有

$$A^{\dagger}=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}1/\sqrt{2} & 0 & 0\\ 0 & 1 & 0\end{bmatrix}\begin{bmatrix}1/\sqrt{2} & 0 & 1/\sqrt{2}\\ 0 & 1 & 0\\ 1/\sqrt{2} & 0 & -1/\sqrt{2}\end{bmatrix}$$

$$=\begin{bmatrix}1/2 & 0 & 1/2\\ 0 & 1 & 0\end{bmatrix}\in \boldsymbol{M}_{2,3}(\mathbb{C}).$$ □

显然，$A^{\dagger}\in A\{1\}$. 下面定理指出，A 的任意$\{1\}$-逆有类似于(3)式的表达式，但其中的 O 应换为大小合适的任意矩阵.

定理 4　设 $A\in \boldsymbol{M}_{m,n}(\mathbb{C})$有秩 $r>0$，且有奇异值分解(1). 则

$$A\{1\}=\{V\begin{bmatrix}\Sigma^{-1} & K\\ L & M\end{bmatrix}U^{*}:L\in \boldsymbol{M}_{n-r,r}(\mathbb{C}),$$
$$K\in \boldsymbol{M}_{r,m-r}(\mathbb{C})\text{ 与 }M\in \boldsymbol{M}_{n-r,m-r}(\mathbb{C})\text{ 任意}\}. \tag{4}$$

证明　直接验证便知，(4)式右端任一矩阵属于 $A\{1\}$. 反之，为证 A 的$\{1\}$-逆具有(4)式右端的形式，显然只要证明：$m\times n$ 矩阵 $B=\begin{bmatrix}\Sigma & O\\ O & O\end{bmatrix}$的$\{1\}$-逆有形式$\begin{bmatrix}\Sigma^{-1} & K\\ L & M\end{bmatrix}$. 现设

$$X=\begin{bmatrix}X_{11} & X_{12}\\ X_{21} & X_{22}\end{bmatrix}\in B\{1\},$$

式中 $X_{11}\in \boldsymbol{M}_r(\mathbb{C})$，则 $BXB=B$ 蕴涵 $X_{11}=\Sigma^{-1}$. □

由定理 4 看出，$A\{1\}$为 $\boldsymbol{M}_{n,m}(\mathbb{C})$中的仿射流形，即 $X,Y\in A\{1\}$与 $\alpha\in\mathbb{C}$ 蕴涵 $\alpha X+(1-\alpha)Y\in A\{1\}$. 此外，$A\{1\}$中矩阵共有 $mn-r^2$ 个独立参数(即 K,L 与 M 元素的个数)，$r=\mathrm{rank}A$. 并且由 $r\leqslant\min\{m,n\}$看出，非零的 $A\in \boldsymbol{M}_{m,n}(\mathbb{C})$有唯一的$\{1\}$-逆当且仅当 $m=n=r>0$，即 A 为非奇异方阵.

对于例 3 中矩阵 A，按(4)式我们有

$$A\{1\}=\left\{\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}1/\sqrt{2} & 0 & \alpha\\ 0 & 1 & \beta\end{bmatrix}\begin{bmatrix}1/\sqrt{2} & 0 & 1/\sqrt{2}\\ 0 & 1 & 0\\ 1/\sqrt{2} & 0 & -1/\sqrt{2}\end{bmatrix},\forall\alpha,\beta\in\mathbb{C}\right\}$$

$$=\left\{\begin{bmatrix}(1+\alpha\sqrt{2})/2 & 0 & (1-\alpha\sqrt{2})/2\\ \beta/\sqrt{2} & 1 & -\beta/\sqrt{2}\end{bmatrix},\forall\alpha,\beta\in\mathbb{C}\right\}. \tag{5}$$

从定义 1 与定理 4 还可以导出 $A^{[1]}$ 的性质.

定理 5　设 $A\in \boldsymbol{M}_{m,n}(\mathbb{C})$ 与 $\alpha\in\mathbb{C}$. 则对任意 $A^{[1]}\in A\{1\}$ 有:

(1) $\operatorname{rank}A\leqslant\operatorname{rank}A^{[1]}\leqslant\min\{m,n\}$.

(2) $(A^{[1]})^*\in A^*\{1\}$, $(A^{[1]})^{\mathrm{T}}\in A^{\mathrm{T}}\{1\}$, $\alpha^\dagger A^{[1]}\in(\alpha A)\{1\}$, 其中,

$$\alpha^\dagger=\begin{cases}0, & 若\ \alpha=0,\\ \alpha^{-1}, & 若\ \alpha\neq 0.\end{cases}\tag{6}$$

(3) 若 $S\in\boldsymbol{M}_m(\mathbb{C})$ 与 $T\in\boldsymbol{M}_n(\mathbb{C})$ 非奇异, 则 $T^{-1}A^{[1]}S^{-1}\in(SAT)\{1\}$.

(4) $AA^{[1]}$ 与 $A^{[1]}A$ 二者均为幂等矩阵, 并且 $\operatorname{Im}A\cap\operatorname{Ker}(A^{[1]})=\{\mathbf{0}\}$; $\operatorname{Im}(A^{[1]}A)\subset\operatorname{Im}A^{[1]}$, $\operatorname{Im}(AA^{[1]})=\operatorname{Im}A$; $\operatorname{Ker}(A^{[1]})\subset\operatorname{Ker}(AA^{[1]})$, $\operatorname{Ker}(A^{[1]}A)=\operatorname{Ker}A$. 此外,

$$\operatorname{rank}(AA^{[1]})=\operatorname{rank}(A^{[1]}A)=\operatorname{rank}A,$$
$$\mathbb{C}^n=\operatorname{Ker}(A^{[1]}A)\dotplus\operatorname{Im}(A^{[1]}A)=\operatorname{Ker}A\dotplus\operatorname{Im}(A^{[1]}A),$$
$$\mathbb{C}^m=\operatorname{Im}(AA^{[1]})\dotplus\operatorname{Ker}(AA^{[1]})=\operatorname{Im}A\dotplus\operatorname{Ker}(AA^{[1]}).$$

(5) A 为列满秩的(即 $\operatorname{rank}A=n$)当且仅当 $A^{[1]}A=I_n$(即 $A^{[1]}$ 为 A 的左逆); A 为行满秩的(即 $\operatorname{rank}A=m$)当且仅当 $AA^{[1]}=I_m$(即 $A^{[1]}$ 为 A 的右逆).

(6) 对 $\mathbb{C}^n$ 内 $\operatorname{Ker}A$ 的任一陪集 $\mathscr{H}$(即 $\boldsymbol{x},\boldsymbol{y}\in\mathscr{H}$ 当且仅当 $\boldsymbol{x}-\boldsymbol{y}\in\operatorname{Ker}A$ 与 $\boldsymbol{x}\in\mathscr{H}$), 交集 $\operatorname{Im}(A^{[1]}A)\cap\mathscr{H}$ 中仅含一个元素 $\boldsymbol{v}$.

证明　若 $\boldsymbol{x}\in\operatorname{Im}A\cap\operatorname{Ker}A^{[1]}$, 则 $A^{[1]}\boldsymbol{x}=\mathbf{0}$ 且对某 $\boldsymbol{y}\in\mathbb{C}^n$ 有 $\boldsymbol{x}=A\boldsymbol{y}$. 于是, $A^{[1]}A\boldsymbol{y}=\mathbf{0}$, 它蕴涵 $AA^{[1]}A\boldsymbol{y}=A\boldsymbol{y}=\boldsymbol{x}=\mathbf{0}$. 因此, $\operatorname{Im}A\cap\operatorname{Ker}A^{[1]}=\{\mathbf{0}\}$. 同理可证: $\operatorname{Im}(A^{[1]}A)\cap\operatorname{Ker}A=\{\mathbf{0}\}$.

现证(6). 设 $\mathscr{H}$ 为 $\operatorname{Ker}A$ 的任一陪集, 且 $\boldsymbol{v}_1,\boldsymbol{v}_2\in\operatorname{Im}(A^{[1]}A)\cap\mathscr{H}$. 则 $\boldsymbol{v}_1-\boldsymbol{v}_2\in\operatorname{Im}(A^{[1]}A)\cap\operatorname{Ker}A=\{\mathbf{0}\}$, 因而 $\boldsymbol{v}_1=\boldsymbol{v}_2$. 现令 $\mathscr{H}=\{\boldsymbol{\alpha}+\operatorname{Ker}A\}$, 其中 $\boldsymbol{\alpha}\in\mathbb{C}^n$ 给定, 又令 $\boldsymbol{v}=A^{[1]}A\boldsymbol{\alpha}$. 则 $\boldsymbol{v}\in\operatorname{Im}(A^{[1]}A)$, 且 $A\boldsymbol{v}=A\boldsymbol{\alpha}+\operatorname{Ker}A$, 这表明: $\boldsymbol{v}\in\operatorname{Im}(A^{[1]}A)\cap\mathscr{H}$. 本定理余下命题由定义 1 或定理 4 直接推出(见习题 4). □

我们指出, 定理 5(4) 中 $\operatorname{Im}(A^{[1]}A)\subset\operatorname{Im}A^{[1]}$ 与 $\operatorname{Ker}A^{[1]}\subset\operatorname{Ker}(AA^{[1]})$ 一般为真正包含关系. 譬如, 令 $A=\operatorname{diag}[1,0\ ,0]$, 则 $A^{[1]}=\operatorname{diag}[1,1\ ,0]\in A\ \{1\}$. 这时, $\operatorname{Im}A^{[1]}=\{(a,b,0)^{\mathrm{T}}:a,b\in\mathbb{C}\}$, $\operatorname{Ker}A^{[1]}=\{(0,0,c)^{\mathrm{T}}:c\in\mathbb{C}\}$, $\operatorname{Im}(A^{[1]}A)=\{(a,0,0)^{\mathrm{T}}:a\in\mathbb{C}\}$, 而 $\operatorname{Ker}A=\operatorname{Ker}(AA^{[1]})=\{(0,b,c)^{\mathrm{T}}:b,c\in\mathbb{C}\}$.

在定理 5(4) 基础之上, 我们可以进一步地得出 $\{1\}$-逆的特征.

定理 6　设 $A\in\boldsymbol{M}_{m,n}(\mathbb{C})$. 则下列诸命题彼此等价:

(1) $X\in A\{1\}$.

(2) AX 为幂等矩阵且 $\operatorname{Im}(AX)=\operatorname{Im}A$.

(3) AX 为幂等矩阵且 $\mathbb{C}^m=\operatorname{Im}A\dotplus\operatorname{Ker}(AX)$.

(4) AX 为幂等矩阵且 $\mathbb{C}^n=\operatorname{Im}X+\operatorname{Ker}A$.

(5) XA 为幂等矩阵且 $\operatorname{Ker}(XA)=\operatorname{Ker}A$.

(6) XA 为幂等矩阵且$\mathbb{C}^n=\mathrm{Ker}A\dotplus\mathrm{Im}(XA)$.

(7) XA 为幂等矩阵且 $\mathrm{Im}A\cap\mathrm{Ker}X=\{\mathbf{0}\}$.

(8) 对任意 $\boldsymbol{b}\in\mathrm{Im}A$,$X\boldsymbol{b}$ 为方程组 $A\boldsymbol{x}=\boldsymbol{b}$ 的一个解,即有 $AX\boldsymbol{b}=\boldsymbol{b}$.

证明 只证(1),(2)与(8)之间的等价性,余下的留给读者.(1)⇒(2):见定理5(4).现设(2),则由 AX 幂等与 $\mathrm{Im}(AX)=\mathrm{Im}A$ 推出,对任意 $\boldsymbol{x}\in\mathbb{C}^n$ 有 $\boldsymbol{y}\in\mathbb{C}^m$ 使得 $A\boldsymbol{x}=AX\boldsymbol{y}$,于是 $AXA\boldsymbol{x}-A\boldsymbol{x}=(AXAX-AX)\boldsymbol{y}=\mathbf{0}$.因此,$AXA=A$.这样一来,(1)⇔(2).(1)⇒(8):若 $\boldsymbol{b}=A\boldsymbol{c}$,则 $AX\boldsymbol{b}=AXA\boldsymbol{c}=A\boldsymbol{c}=\boldsymbol{b}$.(8)⇒(1):令 $A^{(j)}$ 为 A 的第 j 列,$1\leqslant j\leqslant n$.这时,$A\boldsymbol{x}=A^{(j)}$ 有解 $XA^{(j)}$,即 $AXA^{(j)}=A^{(j)}$,$1\leqslant j\leqslant n$,因而 $AXA=A$,即 $X\in A\{1\}$. □

我们指出,定理6(4)中$\mathbb{C}^n=\mathrm{Im}X+\mathrm{Ker}A$ 一般不是直和.例如由定理5后面的例子看到,当 $A=\mathrm{diag}[1,0,0]$,$X=\mathrm{diag}[1,1,0]$时,$\mathrm{Im}X\cap\mathrm{Ker}A\neq\{\mathbf{0}\}$.

由定理6(8)看出,矩阵 A 的$\{1\}$-逆基本应用在于求解相容的线性方程组 $A\boldsymbol{x}=\boldsymbol{b}$.更一般地,它还能给出线性矩阵方程

$$AXB=D \tag{7}$$

的通解的矩阵形式,只要矩阵方程(7)为相容的(即它有解 X).

定理7 设 $A\in\boldsymbol{M}_{m,n}(\mathbb{C})$,$B\in\boldsymbol{M}_{p,q}(\mathbb{C})$与 $D\in\boldsymbol{M}_{m,q}(\mathbb{C})$.则当且仅当对某个 $A^{[1]}\in A\{1\}$与 $B^{[1]}\in B\{1\}$,

$$AA^{[1]}DB^{[1]}B=D \tag{8}$$

时,方程(7)相容.此时,它的通解表达式为

$$X=A^{[1]}DB^{[1]}+W-A^{[1]}AWBB^{[1]},\qquad \forall W\in\boldsymbol{M}_{n,p}(\mathbb{C}) \tag{9}$$

或者

$$X=A^{[1]}DB^{[1]}+(I_n-A^{[1]}A)U+V(I_p-BB^{[1]}),$$
$$\forall U\in\boldsymbol{M}_{n,p}(\mathbb{C}),\qquad V\in\boldsymbol{M}_{n,p}(\mathbb{C}). \tag{9'}$$

证明 若条件(8)成立,则容易看出 $X=A^{[1]}DB^{[1]}$为方程(7)的解.反之,若方程(7)有解 X,那么对任意 $A^{[1]}\in A\{1\}$与 $B^{[1]}\in B\{1\}$,

$$D=AXB=AA^{[1]}AXBB^{[1]}B=AA^{[1]}DB^{[1]}B.$$

当(8)式成立时,直接验证便知(9)式中 X 为方程(7)的解.反之,若 X 为方程(7)解,则 $X=A^{[1]}DB^{[1]}+X-A^{[1]}AXBB^{[1]}$,即 X 有形式(9).在(9)式中取 $W=(I_n-A^{[1]}DB^{[1]}+A)U+V(I_P-BB^{[1]})$即得(9′)式.反之,取 $U=V=\frac{1}{2}(W+A^{[1]}AW+WBB^{[1]})$,(9′)式即为(9)式. □

我们指出,条件(8)等价于

$$AA^{[1]}D=D\quad 与\quad DB^{[1]}B=D. \tag{10}$$

这是因为方程(7)相容性蕴涵 $AY=D$ 与 $XB=D$ 的相容性,后者等价于条件(10).反之,条件(10)显然蕴涵条件(8).

由于 $Ax=b$ 为矩阵方程(7)的特殊情形(取 $p=q=1, B=(1)$)，我们有

推论 8　设 $A\in \boldsymbol{M}_{m,n}(\mathbb{C})$ 与 $\boldsymbol{b}\in \mathbb{C}^m$. 则

(1) 对任意 $A^{[1]}\in A\{1\}$，$\boldsymbol{x}=(I_n-A^{[1]}A)\boldsymbol{y}(\forall \boldsymbol{y}\in \mathbb{C}^n)$ 为 $A\boldsymbol{x}=\boldsymbol{0}$ 的通解.

(2) 方程 $A\boldsymbol{x}=\boldsymbol{b}$ 相容当且仅当对某个 $A^{[1]}\in A\{1\}$，$AA^{[1]}\boldsymbol{b}=\boldsymbol{b}$. 此时 $A\boldsymbol{x}=\boldsymbol{b}$ 的通解表达式为

$$\boldsymbol{x}=A^{[1]}\boldsymbol{b}+(I_n-A^{[1]}A)\boldsymbol{y},\qquad \forall \boldsymbol{y}\in \mathbb{C}^n. \tag{11}$$

(3) 对任意 $\boldsymbol{b}\in \mathbb{C}^m$，$A^{[1]}\boldsymbol{b}$ 为方程 $A\boldsymbol{x}=P_{\mathrm{Im}A,\mathrm{Ker}(AA^{[1]})}\boldsymbol{b}$ 的解.

推论 9　设 $A\in \boldsymbol{M}_{m,n}(\mathbb{C})$ 与 $A^{[1]}\in A\{1\}$. 则

$$A\{1\}=\{A^{[1]}+Z-A^{[1]}AZAA^{[1]},\ \forall Z\in \boldsymbol{M}_{n,m}(\mathbb{C})\}$$

或者

$$A\{1\}=\{A^{[1]}+(I_n-A^{[1]}A)F+G(I_m-AA^{[1]}),\ \forall F,G\in \boldsymbol{M}_{n,m}(\mathbb{C})\}.$$

证明　在(9)式中取 $B=D=A$(因而 $p=m$ 与 $q=n$)，$W=A^{[1]}+Z$，在(9′)式中取 $B=D=A$，$U=F+\frac{1}{2}A^{[1]}$ 与 $V=G+\frac{1}{2}A^{[1]}$. □

从定理 6(3)与(6)看出，若非零 $A\in \boldsymbol{M}_{m,n}(\mathbb{C})$ 有秩 r，它的任一个 $\{1\}$-逆 X 确定出 $\mathbb{C}^n$ 中 $\mathrm{Ker}A$ 的特殊补空间 $\mathscr{T}=\mathrm{Im}(XA)$ 与 $\mathbb{C}^m$ 中 $\mathrm{Im}A$ 的特殊补空间 $\mathscr{S}=\mathrm{Ker}(AX)$ 使得

$$AX=P_{\mathrm{Im}A,\mathscr{S}},\qquad XA=P_{\mathscr{T},\mathrm{Ker}A}=I-P_{\mathrm{Ker}A,\mathscr{T}} \tag{12}$$

反之，自然要问：任意给出 $\mathbb{C}^n$ 中 $\mathrm{Ker}A$ 的补空间 $\mathscr{T}$ 与 $\mathbb{C}^m$ 中 $\mathrm{Im}A$ 的补空间 $\mathscr{S}$，即有

$$\mathbb{C}^n=\mathrm{Ker}A\dotplus\mathscr{T}\quad 与\quad \mathbb{C}^m=\mathrm{Im}A\dotplus\mathscr{S}, \tag{13}$$

(1) 是否存在 $X\in \boldsymbol{M}_{n,m}(\mathbb{C})$ 使得(12)式成立(注意，如这样的 X 存在，则 $AXA=P_{\mathrm{Im}A,\mathscr{S}}A=A$，因而 $X\in A\{1\}$)？

(2) 这些 X 的一般表达式如何？

为此先考虑比方程组(12)更一般的矩阵方程组

$$AX=B,\qquad XD=E \tag{14}$$

的相容性条件以及通解的表达式.

定理 10　设 $A\in \boldsymbol{M}_{m,n}(\mathbb{C})$，$B\in \boldsymbol{M}_{m,p}(\mathbb{C})$，$D\in \boldsymbol{M}_{p,q}(\mathbb{C})$ 与 $E\in \boldsymbol{M}_{n,q}(\mathbb{C})$ 为给定的矩阵.

(1) 矩阵方程组(14)相容当且仅当它的每个矩阵方程相容，且有

$$AE=BD.$$

(2) 若矩阵方程组(14)有解 $X_0\in \boldsymbol{M}_{n,p}(\mathbb{C})$，则其通解为

$$X=X_0+(I_n-A^{[1]}A)Y(I_p-DD^{[1]}),\qquad \forall Y\in \boldsymbol{M}_{n,p}(\mathbb{C}),$$
$$A^{[1]}\in A\{1\},D^{[1]}\in D\{1\}. \tag{15}$$

证明　(1) 必要性是显然的，因为 $AXD=BD=AE$. 反之，对任意 $A^{[1]}\in A\{1\}$ 与 $D^{[1]}\in D\{1\}$，只要 $AE=BD$，$AA^{[1]}B=B$ 与 $ED^{[1]}D=E$，矩阵方程组(14)便有公

共解 X：

$$X = A^{[1]}B + ED^{[1]} - A^{[1]}AED^{[1]}.$$

但按定理 7，条件 $AA^{[1]}B=B$ 与 $ED^{[1]}D=E$ 分别等价于矩阵方程组(14)中两个方程的相容性.

(2) 设 $X_0 \in \boldsymbol{M}_{n,p}(\mathbb{C})$ 为矩阵方程组(14)的公共解. 直接验证便知(15)式中 X 为矩阵方程组(14)的解. 现设 $X \in \boldsymbol{M}_{n,p}(\mathbb{C})$ 为矩阵方程组(14)的任意解，则有

$$X = X_0 + (I_n - A^{[1]}A)(X - X_0)(I_P - DD^{[1]}),$$

因而它有形式(15)(取 $Y=X-X_0$ 即可). □

现返回到矩阵方程组(12). 根据定理 10，方程组(12)相容当且仅当对任意 $A^{[1]} \in A\{1\}$，

$$\begin{aligned} &AA^{[1]}P_{\mathrm{Im}A,\mathscr{S}} = P_{\mathrm{Im}A,\mathscr{S}}, P_{\mathscr{T},\mathrm{Ker}A}A^{[1]}A = P_{\mathscr{T},\mathrm{Ker}A}, \\ &AP_{\mathscr{T},\mathrm{Ker}A} = P_{\mathrm{Im}A,\mathscr{S}}A(= A). \end{aligned} \tag{16}$$

但(16)式中各等式都容易验明. 因此，我们已经证明了下面定理的结果(1).

定理 11 设 $A \in \boldsymbol{M}_{m,n}(\mathbb{C})$，且 $\mathscr{T}$ 与 $\mathscr{S}$ 分别为满足(13)式的$\mathbb{C}^n$ 与$\mathbb{C}^m$ 内任意给定的子空间. 则

(1) 矩阵方程组(12)相容.

(2) $X \in \boldsymbol{M}_{n,m}(\mathbb{C})$ 为矩阵方程组(12)公共解的充分与必要条件是

$$X \in A\{1\}, \mathrm{Ker}(AX) = \mathscr{S}, \qquad \mathrm{Im}(XA) = \mathscr{T}.$$

(3) 对任意给定 $A^{[1]} \in A\{1\}$，矩阵方程组(12)公共通解表达式为

$$X = P_{\mathscr{T},\mathrm{Ker}A}A^{[1]}P_{\mathrm{Im}\,A,\mathscr{S}} + (I_n - A^{[1]}A)Y(I_m - AA^{[1]}), \qquad \forall Y \in \boldsymbol{M}_{n,m}(\mathbb{C}). \tag{17}$$

证明 (2) 由前面知道，X 为矩阵方程组(12)的解蕴涵 $X \in A\{1\}$. 由于 XA 与 AX 均为幂等矩阵，按定理 5，$\mathrm{Im}(XA) = \mathrm{Im}(P_{\mathscr{T},\mathrm{Ker}A}) = \mathscr{T}$，$\mathrm{Ker}(AX) = \mathrm{Ker}(P_{\mathrm{Im}A,\mathscr{S}}) = \mathscr{S}$. 充分性由定理 5 推出，因为当条件成立时，$AX = P_{\mathrm{Im}(AX),\mathrm{Ker}(AX)} = P_{\mathrm{Im}A,\mathscr{S}}$，$XA = P_{\mathrm{Im}(XA),\mathrm{Ker}(XA)} = P_{\mathscr{T},\mathrm{Ker}A}$.

(3) 令 $X_0 = P_{\mathscr{T},\mathrm{Ker}A}A^{[1]}P_{\mathrm{Im}A,\mathscr{S}}$. 则 X_0 为方程组(12)的公共解，这是因为由(16)式，

$$\begin{aligned} AX_0 &= AP_{\mathscr{T},\mathrm{Ker}A}A^{[1]}P_{\mathrm{Im}A,\mathscr{S}} = AA^{[1]}P_{\mathrm{Im}A,\mathscr{S}} = P_{\mathrm{Im}A,\mathscr{S}}, \\ X_0A &= P_{\mathscr{T},\mathrm{Ker}A}A^{[1]}P_{\mathrm{Im}A,\mathscr{S}}A = P_{\mathscr{T},\mathrm{Ker}A}A^{[1]}A = P_{\mathscr{T},\mathrm{Ker}A}. \end{aligned}$$

再应用定理 10(2)便得表达式(17). □

习题 5.1.1

1. 设 $A \in \boldsymbol{M}_{m,n}(\mathbb{C})$有秩 $r>0$. 试证：若存在 $E \in \boldsymbol{M}_m(\mathbb{C})$与 $P \in \boldsymbol{M}_n(\mathbb{C})$使得 E,P 非奇异，且

$$EAP = \begin{bmatrix} I_r & K \\ O & O \end{bmatrix}, \tag{18}$$

则对任意 $L\in \boldsymbol{M}_{n-r,m-r}(\mathbb{C})$,

$$X=P\begin{bmatrix} I_r & O \\ O & L \end{bmatrix}E\in \boldsymbol{M}_{n,m}(\mathbb{C}) \tag{19}$$

属于 $A\{1\}$. 反之,若 $X\in A\{1\}$,则 X 有形式(19),其中 P 与 E 为满足(18)式(但其中 $K=O$)的某两个非奇异矩阵,而 L 为某个 $(n-r)\times(m-r)$ 的复矩阵.

2. 按题 1 结果,求出下列矩阵

$$A=\begin{bmatrix} 0 & 2\mathrm{i} & \mathrm{i} & 0 & 4+2\mathrm{i} & 1 \\ 0 & 0 & 0 & -3 & -6 & -3-3\mathrm{i} \\ 0 & 2 & 1 & 1 & 4-4\mathrm{i} & 1 \end{bmatrix}$$

的形如题 1(19)式中的 $X\in A\{1\}$.

3. 试证:$AXA=A$(这里 $A\in \boldsymbol{M}_{m,n}(\mathbb{C})$有秩 $r>0$)的通解有表达式

$$X=A^{[1]}+FY+BZK,$$

这里,$F\in \boldsymbol{M}_{n,n-r}(\mathbb{C})$, $K^*\in \boldsymbol{M}_{m,m-r}(\mathbb{C})$, $B\in \boldsymbol{M}_{n,r}(\mathbb{C})$为给定的矩阵,它们的列向量分别组成 $\mathrm{Ker}A$, $\mathrm{Ker}A^*$, $\mathrm{Im}(A^{[1]}A)$的基;$A^{[1]}$为 A 的任意给定{1}-逆,$Y\in \boldsymbol{M}_{n-r,m}(\mathbb{C})$与 $Z\in \boldsymbol{M}_{r,m-r}(\mathbb{C})$任意.

4. 证明定理 5(1)~(5).

5. 证明定理 6 中(3)~(7)的等价性以及(2)与(3)之间的等价性.

6. 设 A 如题 2,且 $\boldsymbol{b}=(14+5\mathrm{i},-15+3\mathrm{i},10-15\mathrm{i})^{\mathrm{T}}$. 试写出线性方程组 $A\boldsymbol{x}=\boldsymbol{b}$ 的通解.

7. 设 $A\in \boldsymbol{M}_{m,n}(\mathbb{C})$与 $B\in \boldsymbol{M}_{p,q}(\mathbb{C})$. 试证:

$$A^{[1]}\otimes B^{[1]}\in (A\otimes B)\{1\}.$$

8. 设 $A\in \boldsymbol{M}_{m,n}(\mathbb{C})$与 $X\in A\{1\}$. 试证:$\langle X,Y\rangle=\mathrm{tr}(X^*Y)=0$ 对所有满足 $AYA=O$ 的 $Y\in \boldsymbol{M}_{n,m}(\mathbb{C})$成立,当且仅当 $X=A^{\dagger}$.

9. 设 $A\in \boldsymbol{M}_{m,n}(\mathbb{C})$与 $X\in A\{1\}$. 试证:$\|X+Y\|_F^2=\|X\|_F^2+\|Y\|_F^2$ 对所有满足 $AYA=O$ 的 $Y\in \boldsymbol{M}_{n,n}(\mathbb{C})$成立,当且仅当 $X=A^{\dagger}$. 并且,$A^{\dagger}$ 是 $AXA=A$ 具有最小 Frobenius 范数 $\|\cdot\|_F$ 的解,即若 X 满足 $AXA=A$ 且 $X\neq A^{\dagger}$,则有 $\|A^{\dagger}\|_F<\|X\|_F$.

5.1.2　其他 λ-逆

先考虑{2}-逆与{1,2}-逆.

设 $A\in \boldsymbol{M}_{m,n}(\mathbb{C})$. 显然,$X\in A\{2\}$等价于 $A\in X\{1\}$. 应用这个对称性质,从 5.1.1 小节定理 6 可以直接得到 $A\{2\}$的特征.

定理 1　设 $A\in \boldsymbol{M}_{m,n}(\mathbb{C})$. 则下列诸命题彼此等价:

(1) $X\in A\{2\}$.

(2) XA 为幂等矩阵且 $\mathrm{Im}(XA)=\mathrm{Im}X$.

(3) XA 为幂等矩阵且 $\mathbb{C}^n=\mathrm{Im}X\dot{+}\mathrm{Ker}(XA)$.

(4) XA 为幂等矩阵且 $\mathbb{C}^m=\mathrm{Im}A+\mathrm{Ker}X$.

(5) AX 为幂等矩阵且 $\mathrm{Ker}(AX)=\mathrm{Ker}X$.

(6) AX 为幂等矩阵且 $\mathbb{C}^m=\mathrm{Ker}X\dot{+}\mathrm{Im}(AX)$.

(7) AX 为幂等矩阵且 $\mathrm{Im}X\cap\mathrm{Ker}A=\{\mathbf{0}\}$.

我们指出，由于 $XAX=X$ 不是关于 X 的线性矩阵方程，故 5.1.1 小节中由线性矩阵方程 $AXB=D$ 导出的 $A\{1\}$的特性一般不适于 $A\{2\}$.

对于 $A\{1,2\}$，显然，$X\in A\{1,2\}$蕴涵 $X\in A\{1\}\cap A\{2\}$，并且，$X\in A\{1,2\}$等价于 $A\in X\{1,2\}$，即 A 与 X 互为$\{1,2\}$-逆. 因此，从 5.1.1 小节定理 6 与这里的定理 1 可以直接得到 $A\{1,2\}$的特征.

定理 2 设 $A\in\boldsymbol{M}_{m,n}(\mathbb{C})$与 $X\in\boldsymbol{M}_{n,m}(\mathbb{C})$. 则下列诸命题彼此等价：

(1) $X\in A\{1,2\}$.

(2) AX 为幂等矩阵且$\mathbb{C}^n=\mathrm{Ker}A\dotplus\mathrm{Im}X$.

(3) XA 为幂等矩阵且$\mathbb{C}^m=\mathrm{Im}A\dotplus\mathrm{Ker}X$.

(4) AX 为幂等矩阵且 $\mathrm{Im}(AX)=\mathrm{Im}A$ 与 $\mathrm{Ker}(AX)=\mathrm{Ker}X$.

(5) XA 为幂等矩阵且 $\mathrm{Im}(XA)=\mathrm{Im}X$ 与 $\mathrm{Ker}(XA)=\mathrm{Ker}A$.

由上述定理直接看出，若 $X\in A\{1,2\}$，则这时且仅在这时，AX 为沿着 $\mathrm{Ker}X$ 到 $\mathrm{Im}A$ 上的投影，而 XA 为沿着 $\mathrm{Ker}A$ 到 $\mathrm{Im}X$ 上的投影，即有

$$AX = P_{\mathrm{Im}A,\mathrm{Ker}X},\qquad XA = P_{\mathrm{Im}X,\mathrm{Ker}A}. \tag{1}$$

命题 3 设 $Y,Z\in A\{1\}$. 则 $X=YAZ\in A\{1,2\}$.

证明 直接验证得 $AXA=A$ 与 $XAX=X$. □

显然 $A\{1,2\}\subset A\{1\}$. 下面定理指出 $X\in A\{1\}$为 A 的$\{1,2\}$-逆的充分与必要条件.

定理 4 设 $A\in\boldsymbol{M}_{m,n}(\mathbb{C})$与 $X\in A\{1\}$. 则 $X\in A\{1,2\}$当且仅当 $\mathrm{rank}X=\mathrm{rank}A$.

证明 按定理 2 与 5.1.1 小节定理 5(4)，$X\in A\{1,2\}$蕴涵 XA 幂等与 $\mathrm{rank}(XA)=\mathrm{rank}X=\mathrm{rank}A$. 反之，$X\in A\{1\}$与 $\mathrm{rank}X=\mathrm{rank}A$ 蕴涵 XA 幂等与 $\mathrm{rank}(XA)=\mathrm{rank}A=\mathrm{rank}X$，因而 XA 幂等与 $\mathrm{Im}(XA)=\mathrm{Im}X$，按定理 1 即有 $X\in A\{2\}$. □

推论 5 设 $A\in\boldsymbol{M}_{m,n}(\mathbb{C})$. 则如下三个命题

$$X \in A\{1\}$$
$$X \in A\{2\}$$
$$\mathrm{rank}A = \mathrm{rank}X$$

中任意两个蕴涵余下的命题.

证明 根据定理 4 只需证明：$X\in A\{2\}$与 $\mathrm{rank}A=\mathrm{rank}X$ 蕴涵 $X\in A\{1\}$. 但此只要交换一下 X 与 A 的角色并应用定理 4 即可推得. □

推论 6 设 $A\in\boldsymbol{M}_{m,n}(\mathbb{C})$，且 $\mathscr{T}$ 与 $\mathscr{S}$ 同 5.1.1 小节定理 11. 令 $A_{\mathscr{T},\mathscr{S}}^{[1,2]}=P_{\mathscr{T},\mathrm{Ker}A}A^{[1]}P_{\mathrm{Im}A,\mathscr{S}}$，其中 $A^{[1]}\in A\{1\}$. 则 $A_{\mathscr{T},\mathscr{S}}^{[1,2]}$为 $A\{1,2\}$中满足条件

$$\mathrm{Im}A_{\mathscr{T},\mathscr{S}}^{[1,2]} = \mathscr{T},\qquad \mathrm{Ker}A_{\mathscr{T},\mathscr{S}}^{[1,2]} = \mathscr{S} \tag{2}$$

的唯一元素.

证明　由 5.1.1 小节定理 11(3)证明知道,$A^{[1,2]}_{\mathcal{T},\mathcal{S}}$为 5.1.1 小节方程组(12)的解,因而它属于 $A\{1\}$. 现证明 $A^{[1,2]}_{\mathcal{T},\mathcal{S}}\in A\{1,2\}$. 由于 $A^{[1,2]}_{\mathcal{T},\mathcal{S}}\in A\{1\}$,故按 5.1.1 小节定理 5(4)我们有 $\operatorname{rank}(AA^{[1,2]}_{\mathcal{T},\mathcal{S}})=\operatorname{rank}A=\operatorname{rank}(A^{[1,2]}_{\mathcal{T},\mathcal{S}}A)$. 因此,

$$\operatorname{rank}(P_{\mathcal{T},\operatorname{Ker}A}A^{[1]}P_{\operatorname{Im}A,\mathcal{S}})\leqslant\operatorname{rank}P_{\operatorname{Im}A,\mathcal{S}}=\operatorname{rank}(AA^{[1,2]}_{\mathcal{T},\mathcal{S}})=\operatorname{rank}A\leqslant\operatorname{rank}A^{[1,2]}_{\mathcal{T},\mathcal{S}},$$

即有 $\operatorname{rank}A^{[1,2]}_{\mathcal{T},\mathcal{S}}=\operatorname{rank}A$. 应用定理 4 我们有 $A^{[1,2]}_{\mathcal{T},\mathcal{S}}\in A\{1,2\}$. 并且,从定理 2(4)与(5)以及 5.1.1 小节定理 11(2)看出(2)式成立.

现证明 $A^{[1,2]}_{\mathcal{T},\mathcal{S}}$的唯一性. 设 $X\in A\{1,2\}$满足 $\operatorname{Im}X=\mathcal{T}$与 $\operatorname{Ker}X=\mathcal{S}$.按(1)式,

$$AX=P_{\operatorname{Im}A,\mathcal{S}},\qquad XA=P_{\mathcal{T},\operatorname{Ker}A}.$$

由此得出

$$\begin{aligned}X-A^{[1,2]}_{\mathcal{T},\mathcal{S}}&=XA(X-A^{[1,2]}_{\mathcal{T},\mathcal{S}})+(X-A^{[1,2]}_{\mathcal{T},\mathcal{S}})AA^{[1,2]}_{\mathcal{T},\mathcal{S}}\\&=X(P_{\operatorname{Im}A,\mathcal{S}}-P_{\operatorname{Im}A,\mathcal{S}})+(P_{\mathcal{T},\operatorname{Ker}A}-P_{\mathcal{T},\operatorname{Ker}A})A^{[1,2]}_{\mathcal{T},\mathcal{S}}=O,\end{aligned}$$

即有 $X=A^{[1,2]}_{\mathcal{T},\mathcal{S}}$. □

应用 5.1.1 小节定理 4 与这里的定理 4,我们可以给出 $A\{1,2\}$的结构特征.

定理 7　设 $A\in\boldsymbol{M}_{m,n}(\mathbb{C})$有秩 $r>0$,且它有奇异值分解如 5.1.1 小节(1)式. 则

$$A\{1,2\}=\left\{V\begin{bmatrix}\Sigma^{-1}&K\\L&L\Sigma K\end{bmatrix}U^*:L,K\text{ 任意}\right\}.\tag{3}$$

证明　设 $X\in A\{1,2\}$,按 5.1.1 小节定理 4 与这里的定理 4,它等价于:有 K, L 与 M 使得

$$X=V\begin{bmatrix}\Sigma^{-1}&K\\L&M\end{bmatrix}U^*,\qquad\operatorname{rank}X=r.$$

但是,

$$\begin{bmatrix}I_r&0\\-L\Sigma&I_{n-r}\end{bmatrix}\begin{bmatrix}\Sigma^{-1}&K\\L&M\end{bmatrix}=\begin{bmatrix}\Sigma^{-1}&K\\0&M-L\Sigma K\end{bmatrix},$$

因而 $M=L\Sigma K$(否则 $\operatorname{rank}X>r$). □

其次,考虑 $A\{1,3\}$与 $A\{1,4\}$.

若 $A\in\boldsymbol{M}_{m,n}(\mathbb{C})$,且 $X\in A\{1,3\}$,则 AX 为 Hermite 幂等矩阵,因而 AX 为 $\operatorname{Im}(AX)=\operatorname{Im}A$ 上的正交投影 $P_{\operatorname{Im}A}$,反之亦然. 由此得到

$$X\in A\{1,3\}\Leftrightarrow AX=P_{\operatorname{Im}A}.\tag{4}$$

由于当 $X\in A\{1,4\}$时,$\operatorname{Im}(XA)=\operatorname{Im}(A^*X^*)=\operatorname{Im}A^*=(\operatorname{Ker}A)^{\perp}$,故同理可得

$$X\in A\{1,4\}\Leftrightarrow XA=P_{(\operatorname{Ker}A)^{\perp}}=P_{\operatorname{Im}A^*}.\tag{5}$$

另一方面,从 5.1.1 小节定理 4 中 $A\{1\}$表达式看出,为使得 $X\in A\{1\}$满足 $AX=(AX)^*$,当且仅当其中的 K 为零矩阵. 因此,若 A 的奇异值分解为 5.1.1 小

节(1)式,那么,

$$A\{1,3\} = \left\{V\begin{bmatrix}\Sigma^{-1} & O\\ L & M\end{bmatrix}U^* : L, M \text{ 任意}\right\}. \tag{6}$$

同理可证,

$$A\{1,4\} = \left\{V\begin{bmatrix}\Sigma^{-1} & K\\ O & M\end{bmatrix}U^* : K, M \text{ 任意}\right\}. \tag{7}$$

从(6)或(4)式容易看出,对任意给定 $A^{[1,3]}\in A\{1,3\}$,集合 $A\{1,3\}$ 恰为矩阵方程

$$AX = AA^{[1,3]} \tag{8}$$

的解集. 同理从(7)或(5)式看出,对任意给定 $A^{[1,4]}\in A\{1,4\}$,集合 $A\{1,4\}$恰为矩阵方程

$$XA = A^{[1,4]}A \tag{9}$$

的解集. 应用 5.1.1 小节(9′)式我们有

$$A\{1,3\} = \{A^{[1,3]} + (I_n - A^{[1,3]}A)Z : Z \in \boldsymbol{M}_{n,m}(\mathbb{C})\}, \tag{10}$$

$$A\{1,4\} = \{A^{[1,4]} + Y(I_m - AA^{[1,4]}) : Y \in \boldsymbol{M}_{n,m}(\mathbb{C})\}. \tag{11}$$

综上所述,我们得到

定理 8　设 $A\in \boldsymbol{M}_{m,n}(\mathbb{C})$有奇异值分解如 5.1.1 小节(1)式,且 $A^{[1,3]}\in A\{1,3\}$ 与 $A^{[1,4]}\in A\{1,4\}$任意给定. 则

(1) $A\{1,3\}$有表达式如(6)或(10)式.

(2) $A\{1,4\}$有表达式如(7)或(11)式.

对于 $A\{1,2,3\}$与 $A\{1,2,4\}$,作为定理 8 的推论,我们有(见习题 7)

推论 9　设 $A\in \boldsymbol{M}_{m,n}(\mathbb{C})$有奇异值分解如 5.1.1 小节(1)式,并且 $A^{[1,2,3]}\in A\{1,2,3\}$与 $A^{[1,2,4]}\in A\{1,2,4\}$任意给定. 则

(1)

$$A\{1,2,3\} = \left\{V\begin{bmatrix}\Sigma^{-1} & O\\ L & O\end{bmatrix}U^* : L \text{ 任意}\right\}, \tag{12}$$

$$A\{1,2,3\} = \{A^{[1,2,3]} + (I_n - A^{[1,2,3]}A)ZA^{[1,2,3]} : Z \in \boldsymbol{M}_n(\mathbb{C})\}. \tag{13}$$

(2)

$$A\{1,2,4\} = \left\{V\begin{bmatrix}\Sigma^{-1} & K\\ O & O\end{bmatrix}U^* : K \text{ 任意}\right\}, \tag{14}$$

$$A\{1,2,4\} = \{A^{[1,2,4]} + A^{[1,2,4]}Z(I_m - AA^{[1,2,4]}) : Z \in \boldsymbol{M}_m(\mathbb{C})\}. \tag{15}$$

推论 10　设 $A\in \boldsymbol{M}_{m,n}(\mathbb{C})$. 则

(1) $Y=(A^*A)^{[1]}A^* \in A\{1,2,3\}$.

(2) $Z=A^*(AA^*)^{[1]}\in A\{1,2,4\}$.

例如,对于 5.1.1 小节例 3 中矩阵 A,按推论 9,

$$A\{1,2,3\}=\left\{\begin{bmatrix}1&0\\0&1\end{bmatrix}\begin{bmatrix}1/\sqrt{2}&0&0\\0&1&0\end{bmatrix}\begin{bmatrix}1/\sqrt{2}&0&1/\sqrt{2}\\0&1&0\\1/\sqrt{2}&1&-1/\sqrt{2}\end{bmatrix}\right\}$$

$$=\left\{\begin{bmatrix}1/2&0&1/2\\0&1&0\end{bmatrix}\right\}=\{A^\dagger\}.$$

$$A\{1,2,4\}=\left\{\begin{bmatrix}(1+\alpha\sqrt{2})/2&0&(1-\alpha\sqrt{2})/2)\\\beta/\sqrt{2}&1&-\beta/\sqrt{2}\end{bmatrix}:\forall\alpha,\beta\in\mathbb{C}\right\}=A\{1\}.$$

我们指出，由于 $A\{1,2,3,4\}=A\{1,2,3\}\cap A\{1,2,4\}$，故从推论 9 也可证明 5.1.1 小节(3)式成立.

利用 $A\in\boldsymbol{M}_{m,n}(\mathbb{C})$ 的 $\{1,3\}$-逆与 $\{1,4\}$-逆的表达式(6)与(7)，可以给出 $A^\dagger$ 的另一种表达式.

推论 11　设 $A\in\boldsymbol{M}_{m,n}(\mathbb{C})$，且 $A^{[1,4]}\in A\{1,4\}$ 与 $A^{[1,3]}\in A\{1,3\}$ 任意给定. 则

$$A^\dagger=A^{[1,4]}AA^{[1,3]}.\tag{16}$$

$A^\dagger$ 的其他几种显式表示(见后面(20)式)得自矩阵 A 的**满秩分解**.

定理 12(矩阵的满秩分解定理)　设 $A\in\boldsymbol{M}_{m,n}(\mathbb{C})$ 有秩 $r>0$. 则存在 $F\in\boldsymbol{M}_{m,r}(\mathbb{C})$，其秩为 r，与 $G\in\boldsymbol{M}_{r,n}(\mathbb{C})$，其秩为 r，使得

$$A=FG.\tag{17}$$

证明　取 $F\in\boldsymbol{M}_{m,r}(\mathbb{C})$ 使其列向量恰为 $\mathrm{Im}A$ 的基. 这时，$\mathrm{rank}F=r$，且 A 的每一列向量可唯一地表示为 F 的列向量的线性组合，因而存在唯一的 $G\in\boldsymbol{M}_{r,m}(\mathbb{C})$ 使得(17)式成立. 由于 $r\geqslant\mathrm{rank}G\geqslant\mathrm{rank}A=r$，故有 $\mathrm{rank}G=r$. □

我们指出，(17)式中 A 的满秩分解不是唯一的. 因为对任意非奇异 $Q\in\boldsymbol{M}_r(\mathbb{C})$，令 $F_1=FQ$ 与 $G_1=Q^{-1}G$，则有 $A=F_1Q_1$，它仍为 A 的满秩分解.

若已知 $A\in\boldsymbol{M}_{m,n}(\mathbb{C})(\mathrm{rank}A=r)$ 的(在等价意义下)最简单法式

$$A=P\begin{bmatrix}I_r&O\\O&O\end{bmatrix}_{m\times n}Q,\tag{18}$$

其中，$P\in\boldsymbol{M}_m(\mathbb{C})$ 与 $Q\in\boldsymbol{M}_n(\mathbb{C})$ 非奇异，令 $P=[\underset{r}{F}\ \ \underset{m-r}{F_1}]$ 与 $Q=\begin{bmatrix}G\\G_1\end{bmatrix}\begin{matrix}r\\n-r\end{matrix}$，则 $F\in\boldsymbol{M}_{m,r}(\mathbb{C})$ 列满秩，$G\in\boldsymbol{M}_{r,n}(\mathbb{C})$ 行满秩，并且，A 有满秩分解 $A=FG$.

例如，对 5.1.1 小节例 3 中矩阵 A，(18)式为

$$A=\begin{bmatrix}1/\sqrt{2}&0&1/\sqrt{2}\\0&1&0\\1/\sqrt{2}&0&-1\sqrt{2}\end{bmatrix}\begin{bmatrix}1&0\\0&1\\0&0\end{bmatrix}\begin{bmatrix}\sqrt{2}&0\\0&1\end{bmatrix}.$$

于是，

$$F=\begin{bmatrix}1/\sqrt{2} & 0\\ 0 & 1\\ 1/\sqrt{2} & 0\end{bmatrix},\quad G=\begin{bmatrix}\sqrt{2} & 0\\ 0 & 1\end{bmatrix},\quad A=FG. \tag{19}$$

定理 13 设 $A\in \boldsymbol{M}_{m,n}(\mathbb{C})$ 有秩 $r>0$，且其满秩分解为(17)式. 则有

$$\begin{gathered}G^{[i]}F^{[1]}\in A\{i\},i=1,2,4,\\ G^{[1]}F^{[j]}\in A\{j\},j=1,2,3,\\ A^{\dagger}=G^{*}(F^{*}AG^{*})^{-1}F^{*}=G^{\dagger}F^{[1,3]}=G^{[1,4]}F^{\dagger}=G^{\dagger}F^{\dagger}.\end{gathered} \tag{20}$$

证明 按 5.1.1 小节定理 5(5)，$F^{[1]}F=I_r$ 与 $GG^{[1]}=I_r$. 因此，$G^{[1]}F^{[1]}\in A\{1\}$. 对 $i=2$，$G^{[2]}F^{[1]}FGG^{[2]}F^{[1]}=G^{[2]}GG^{[2]}F^{[1]}=G^{[2]}F^{[1]}$，这表明 $G^{[2]}F^{[1]}\in A\{2\}$. 对 $i=4$，$G^{[4]}F^{[1]}FG=G^{[4]}G=(G^{[4]}G)^{*}$，于是 $G^{[4]}F^{[1]}FG$ 是 Hermite 矩阵，亦即：$G^{[4]}F^{[1]}\in A\{4\}$. 类似可证 $G^{[1]}F^{[j]}\in A\{j\}$，$j=1,2,3$. 由这些事实立即得出，$A^{\dagger}=G^{\dagger}F^{[1,3]}=G^{[1,4]}F^{\dagger}$ 与 $A^{\dagger}=G^{\dagger}F^{\dagger}$. 余下证明：$A^{\dagger}=G^{*}(F^{*}AG^{*})^{-1}F^{*}$. 先证 $F^{*}AG^{*}\in \boldsymbol{M}_r(\mathbb{C})$ 非奇异. 事实上，$F^{*}AG^{*}=(F^{*}F)(GG^{*})$，而 $F^{*}F$ 与 GG^{*} 皆为 r 阶非奇异矩阵. 这时，$G^{*}(F^{*}AG^{*})^{-1}F^{*}=G^{*}(GG^{*})^{-1}(F^{*}F)^{-1}F^{*}$，记为 X，直接验证知道(或按推论 10)，$G^{*}(GG^{*})^{-1}=G^{\dagger}$ 与 $(F^{*}F)^{-1}F^{*}=F^{\dagger}$. 因此，$X=G^{\dagger}F^{\dagger}=A^{\dagger}$. □

特殊地，若 $F\in \boldsymbol{M}_{m,r}(\mathbb{C})$ 列满秩，则由定理 13 证明得

$$F^{\dagger}=(F^{*}F)^{-1}F^{*}. \tag{21}$$

同理，若 $G\in \boldsymbol{M}_{r,n}(\mathbb{C})$ 行满秩，则有

$$G^{\dagger}=G^{*}(GG^{*})^{-1}. \tag{22}$$

例 14 求矩阵

$$A=\begin{bmatrix}0 & 1 & 2 & -1\\ -2 & 0 & 0 & 6\\ 4 & -2 & -4 & -10\end{bmatrix}$$

的 $A^{\dagger}$. 经过计算得到 A 的形如(18)的表达式为

$$A=\begin{bmatrix}0 & 1 & 0\\ -2 & 0 & 0\\ 4 & -2 & 1\end{bmatrix}\begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 0\end{bmatrix}\begin{bmatrix}1 & 0 & 0 & -3\\ 0 & 1 & 2 & -1\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{bmatrix}.$$

由于 $\mathrm{rank}A=2$，故有

$$F=\begin{bmatrix}0 & 1\\ -2 & 0\\ 4 & -2\end{bmatrix} \text{与} G=\begin{bmatrix}1 & 0 & 0 & -3\\ 0 & 1 & 2 & -1\end{bmatrix},$$

因而由 $A^{\dagger}=G^{*}(F^{*}AG^{*})^{-1}F^{*}$ 得到

$$A^{\dagger}=\begin{bmatrix}1 & 0\\0 & 1\\0 & 2\\-3 & -1\end{bmatrix}\frac{1}{1836}\begin{bmatrix}6 & -12\\65 & 176\end{bmatrix}\begin{bmatrix}0 & -2 & 4\\1 & 0 & -2\end{bmatrix}$$

$$=\frac{1}{1836}\begin{bmatrix}-12 & -12 & 48\\176 & -130 & -92\\352 & -260 & -184\\-140 & 166 & -52\end{bmatrix}.$$

□

矩阵 A 的 Moore-Penrose 逆 $A^{\dagger}$ 有许多重要性质，其中一部分类似于非奇异矩阵的逆，但后者的许多性质对 $A^{\dagger}$ 已不复存在.

定理 15 设 $A\in\boldsymbol{M}_{m,n}(\mathbb{C})$. 则 $A^{\dagger}$ 有如下性质：

(1) $(A^{\dagger})^{\dagger}=A,(A^{*})^{\dagger}=(A^{\dagger})^{*},(A^{\mathrm{T}})^{\dagger}=(A^{\dagger})^{\mathrm{T}}$.

(2) $\operatorname{rank}A=\operatorname{rank}A^{\dagger}=\operatorname{rank}(A^{\dagger}A)=\operatorname{rank}(AA^{\dagger})$.

(3) $(AA^{*})^{\dagger}=(A^{*})^{\dagger}A^{\dagger},(A^{*}A)^{\dagger}=A^{\dagger}(A^{*})^{\dagger}$.

(4) $(AA^{*})^{\dagger}AA^{*}=AA^{\dagger},(A^{*}A)^{\dagger}A^{*}A=A^{\dagger}A$.

(5) $A^{\dagger}=(A^{*}A)^{\dagger}A^{*}=A^{*}(AA^{*})^{\dagger}$. 特别地，若 A 列满秩，则 $A^{\dagger}=(A^{*}A)^{-1}A^{*}$；若 A 行满秩，则 $A^{\dagger}=A^{*}(AA^{*})^{-1}$.

(6) $(UAV)^{\dagger}=V^{*}A^{\dagger}U^{*},\forall U\in\mathscr{U}_m,V\in\mathscr{U}_n$.

(7) 若 $A=\begin{bmatrix}D & O\\O & O\end{bmatrix},D=\operatorname{diag}[d_1,d_2,\cdots,d_r],d_i\neq0,1\leqslant i\leqslant r$，则 $\operatorname{rank}A^{\dagger}=r$，且 $A^{\dagger}=\begin{bmatrix}D^{-1} & O\\O & O\end{bmatrix}\in\boldsymbol{M}_{n,m}(\mathbb{C})$.

(8) $(A\otimes B)^{\dagger}=A^{\dagger}\otimes B^{\dagger}$，这里 $B\in\boldsymbol{M}_{p,q}(\mathbb{C})$.

本定理的证明留给读者(见习题 9).

应用定理 15(2)与(5)容易推出 $\operatorname{Im}A^{*}=\operatorname{Im}A^{\dagger}=\operatorname{Im}(A^{\dagger}A)$. 于是，由(4)与(5)式看出，

$$AA^{\dagger}=P_{\operatorname{Im}A},\qquad A^{\dagger}A=P_{\operatorname{Im}A^{\dagger}},\tag{23}$$

即 $AA^{\dagger}$ 为 $\operatorname{Im}A$ 上的正交投影，$A^{\dagger}A$ 为 $\operatorname{Im}A^{*}=\operatorname{Im}A^{\dagger}$ 上的正交投影. 并且，$X=A^{\dagger}$ 为线性矩阵方程组

$$AX=P_{\operatorname{Im}A},\qquad XA=P_{\operatorname{Im}X}\tag{24}$$

的唯一解，这是因为方程组(24)直接蕴涵四个 Penrose 方程，反之亦然. 这表明 Moore 意义下的矩阵广义逆恰为 Moore-Penrose 逆(见本章的前言部分). 根据推论 6 我们顺便地得到

$$A^{\dagger}=A^{[1,2]}_{\operatorname{Im}A^{*},\operatorname{Ker}A^{*}}=P_{\operatorname{Im}A^{*}}A^{[1]}P_{\operatorname{Im}A}.\tag{25}$$

因此我们有如下定理.

定理 16 设$A\in \boldsymbol{M}_{m,n}(\mathbb{C})$. 则$A^{\dagger}$为$A\{1,2\}$中满足条件:$\mathrm{Im}X=\mathrm{Im}A^{*}$与$\mathrm{Ker}X=\mathrm{Ker}A^{*}$的唯一元素. 并且,Penrose 方程全体等价于矩阵方程组(24).

习题 5.1.2

1. 对习题 5.1.1 题 2 中矩阵A求它的$\{1,2\}$-逆,$\{1,2,3\}$-逆,$\{1,2,4\}$-逆与$A^{\dagger}$.

2. 设$A\in \boldsymbol{M}_{m,n}(\mathbb{C})$有秩$r>0$,且$0<s\leqslant r$. 试证:$A\{2\}$中秩为$s$的矩阵集合为

$$A\{2\}_s=\{YZ:Y\in \boldsymbol{M}_{n,s}(\mathbb{C}),Z\in \boldsymbol{M}_{s,m}(\mathbb{C}),ZAY=I_s\}.$$

应用这个结果与推论 5 证明:

$$A\{1,2\}=\{YZ:Y\in \boldsymbol{M}_{n,r}(\mathbb{C}),Z\in \boldsymbol{M}_{r,m}(\mathbb{C}),ZAY=I_r\}.$$

3. 设$A\in \boldsymbol{M}_{m,n}(\mathbb{C})$有秩$r>0$,且$0<s\leqslant r$. 试证:

$$A\{2,3\}_s=\{Y(AY)^{\dagger}:AY\in \boldsymbol{M}_{m,s}(\mathbb{C})\text{ 有秩 }s\}$$

为$A\{2,3\}$中秩为s的矩阵集合.

4. 设$A\in \boldsymbol{M}_{m,n}(\mathbb{C})$有秩$r>0$. 试证:

$$A\{1,2,3\}=\{Y(AY)^{\dagger}:AY\in \boldsymbol{M}_{m,r}(\mathbb{C})\text{ 有秩 }r\},$$

$$A\{1,2,4\}=\{(YA)^{\dagger}Y:YA\in \boldsymbol{M}_{r,n}(\mathbb{C})\text{ 有秩 }r\}.$$

5. 设$A\in \boldsymbol{M}_{m,n}(\mathbb{C})$有秩$r>0$,且$U\in \boldsymbol{M}_{n,p}(\mathbb{C})$与$V\in \boldsymbol{M}_{q,m}(\mathbb{C})$. 令$X=U(VAU)^{[1]}V$,这里$(VAU)^{[1]}$为$(VAU)\{1\}$中任一给定的元素. 试证:

(1) $X\in A\{1\}$当且仅当$\mathrm{rank}(VAU)=r$.

(2) $X\in A\{2\}$与$\mathrm{Im}X=\mathrm{Im}U$当且仅当$\mathrm{rank}(VAU)=\mathrm{rank}U$.

(3) $X\in A\{2\}$与$\mathrm{Ker}X=\mathrm{Ker}V$当且仅当$\mathrm{rank}(VAU)=\mathrm{rank}V$.

(4) $X=A^{[1,2]}_{\mathrm{Im}U,\mathrm{Ker}V}$当且仅当$\mathrm{rank}U=\mathrm{rank}V=\mathrm{rank}(VAU)=r$.

6. 设A为习题 5.1.1 题 2 中的矩阵,且$\mathscr{S}$为由$(0,0,1)^{\mathrm{T}}$生成的$\mathbb{C}^3$子空间,$\mathscr{T}$为由$(0,0,1,0,0,0)^{\mathrm{T}}$与$(0,0,0,1,0,0)^{\mathrm{T}}$生成的$\mathbb{C}^6$子空间. 试求$A^{[1,2]}_{\mathscr{T},\mathscr{S}}$.

7. 证明推论 9,10 与 11.

8. 证明定理 13(20)式中第二个结果:$G^{[1]}F^{[j]}\in A\{j\},j=1,2,3$.

9. 证明定理 15.

10. 设$A\in \boldsymbol{M}_{m,n}(\mathbb{C})$与$X\in \boldsymbol{M}_{n,m}(\mathbb{C})$. 试证:$X\in A\{2\}$当且仅当它有形式$X=(EAF)^{\dagger}$,这里$E$与$F$皆为 Hermite 幂等矩阵.

11. 试证:对任意$A\in \boldsymbol{M}_{m,n}(\mathbb{C})$,$A^{\dagger}$为矩阵方程组

$$AX=P_{\mathrm{Im}A},\qquad XA=P_{\mathrm{Im}A^{*}},\qquad XAX=X$$

的唯一公共解X.

12. 试证:对任意$A\in \boldsymbol{M}_{m,n}(\mathbb{C})$,

$$A\{2,3,4\}=\{YA^{*}:Y=Y^{*}\in (A^{*}A)\{2,4\}\}.$$

并且,$A\{2,3,4\}$为有限集合当且仅当$A^{*}A$的非零特征值互不相同. 同时,若$A^{*}A$只有k个不同的非零特征值,则$A\{2,3,4\}$恰有2^k元素.

13. 试证:矩阵

$$A=\frac{1}{10}\begin{bmatrix}9-3i & 12-4i & 10-10i\\ 3-3i & 4-4i & 0\\ 6+6i & 8+8i & 0\\ 6 & 8 & 0\end{bmatrix}$$

恰有四个{2,3,4}-逆,即

$$X_1=A^{\dagger}=\frac{1}{70}\begin{bmatrix}0 & 6+6i & 12-12i & 12\\ 0 & 8+8i & 16-16i & 16\\ 35+35i & -5-15i & -30+10i & -20-10i\end{bmatrix},$$

$$X_2=\frac{1}{60}\begin{bmatrix}-9-3i & 3+3i & 6-6i & 6\\ -12-4i & 4+4i & 8-8i & 8\\ 25+25i & -5-15i & -30+10i & -20-10i\end{bmatrix},$$

$$X_3=\frac{1}{420}\begin{bmatrix}63+21i & 15+15i & 30-30i & 30\\ 84+28i & 20+20i & 40-40i & 40\\ 35+35i & 5+15i & 30-10i & 20+10i\end{bmatrix},$$

$$X_4=O.$$

14. 设 $A\in\boldsymbol{M}_{m,n}(\mathbb{C})$ 有秩 r,且 $r<s\leqslant\min\{m,n\}$. 试证:$A\{1\}$ 中秩为 s 的矩阵集合为

$$A\{1\}_s=\left\{YZ:Y\in\boldsymbol{M}_{n,s}(\mathbb{C}),Z\in\boldsymbol{M}_{s,m}(\mathbb{C})\text{ 皆有秩 }s,ZAY=\begin{bmatrix}I_r & O\\ O & O\end{bmatrix}\right\}.$$

5.1.3 在求解线性矩阵方程问题中的应用

由前面知道,矩阵的广义逆与所谓"不适定"的矩阵方程(组)的求解问题有着密切联系. 在那里矩阵 $A\in\boldsymbol{M}_{m,n}(\mathbb{C})$ 的{1}-逆,{1,2}-逆,{1,3}-逆,{1,4}-逆与 $A^{\dagger}$ 在求解相容的线性矩阵方程(组)中起着基本作用. 本小节继续研究它们在求解相容或不相容的线性矩阵方程问题中的应用.

定理 1 设 $A\in\boldsymbol{M}_{m,n}(\mathbb{C})$ 与 $\boldsymbol{b}\in\mathbb{C}^m$. 若 $A\boldsymbol{x}=\boldsymbol{b}$ 相容,则对任意 $X\in A\{1,4\}$,$\boldsymbol{x}=X\boldsymbol{b}$ 给出 $A\boldsymbol{x}=\boldsymbol{b}$ 具有最小欧氏范数 $\|\cdot\|_2$ 的唯一解.

证明 若 $\boldsymbol{b}\in\mathrm{Im}A$ 与 $X\in A\{1,4\}$,则按 5.1.1 小节定理 6(8),$A(X\boldsymbol{b})=\boldsymbol{b}$. 现设 $\boldsymbol{x}_0$ 为 $A\boldsymbol{x}=\boldsymbol{b}$ 的任一解. 由 5.1.1 小节推论 8,有 $\boldsymbol{y}\in\mathbb{C}^n$ 使得

$$\boldsymbol{x}_0=X\boldsymbol{b}+(I_n-XA)\boldsymbol{y}. \tag{1}$$

另一方面,按 5.1.2 小节(5)式,$X\boldsymbol{b}\in\mathrm{Im}(XA)=\mathrm{Im}A^*=(\mathrm{Ker}A)^{\perp}$,因而 $X\boldsymbol{b}$ 与 $(I_n-XA)\boldsymbol{y}\in\mathrm{Ker}A$ 正交:$\langle X\boldsymbol{b},(I_n-XA)\boldsymbol{y}\rangle=0$. 于是,

$$\|\boldsymbol{x}_0\|_2^2=\|X\boldsymbol{b}\|_2^2+\|(I_n-XA)\boldsymbol{y}\|_2^2.$$

由此看出,若 $\boldsymbol{x}_0\neq X\boldsymbol{b}$,则 $(I_n-XA)\boldsymbol{y}\neq\boldsymbol{0}$,因而 $\|X\boldsymbol{b}\|_2<\|\boldsymbol{x}_0\|_2$. □

下面推论告诉我们,定理 1 的逆命题也真.

推论 2 设 $A\in\boldsymbol{M}_{m,n}(\mathbb{C})$. 则对任意 $\boldsymbol{b}\in\mathrm{Im}A$,$\boldsymbol{x}=X\boldsymbol{b}$ 为 $A\boldsymbol{x}=\boldsymbol{b}$ 最小欧氏范数解的充分与必要条件为 $X\in A\{1,4\}$.

证明　充分性证明见定理 1. 现设 $X\in \boldsymbol{M}_{n,m}(\mathbb{C})$使得 $X\boldsymbol{b}$ 为 $A\boldsymbol{x}=\boldsymbol{b}$ 的最小欧氏范数解，$\forall\, \boldsymbol{b}\in \mathrm{Im}A$. 取 $\boldsymbol{b}$ 为 A 的第 j 列 $A^{(j)}$，$j=1,\cdots,n$，按定理 1，$A^{[1,4]}A^{(j)}$ 给出 $A\boldsymbol{x}=A^{(j)}$ 的具有最小欧氏范数 $\|\cdot\|_2$ 的唯一解，因而 $XA^{(j)}=A^{[1,4]}A^{(j)}$，$j=1,\cdots,n$，此等价于 $XA=A^{[1,4]}A$. 由 5.1.2 小节(9)式知道，这时有 $X\in A\{1,4\}$. □

假如线性方程组 $A\boldsymbol{x}=\boldsymbol{b}$ 不相容(即 $\boldsymbol{b}\notin \mathrm{Im}A$)，我们可以考虑使得

$$\|A\boldsymbol{x}-\boldsymbol{b}\|_2=\min \tag{2}$$

的所谓**最小二乘解** $\boldsymbol{x}\in\mathbb{C}^n$. 应用 A 的$\{1,3\}$-逆，有如下基本结果.

定理 3　设 $A\in \boldsymbol{M}_{m,n}(\mathbb{C})$与 $\boldsymbol{b}\in\mathbb{C}^m$. 则对任意 $X\in A\{1,3\}$，$\boldsymbol{x}=X\boldsymbol{b}$ 为 $A\boldsymbol{x}=\boldsymbol{b}$ 的最小二乘解. 并且，$A^{\dagger}\boldsymbol{b}$ 为 $A\boldsymbol{x}=\boldsymbol{b}$ 的具有最小欧氏范数的唯一最小二乘解.

证明　将 $A\boldsymbol{x}-\boldsymbol{b}$ 改写为

$$A\boldsymbol{x}-\boldsymbol{b}=(A\boldsymbol{x}-P_{\mathrm{Im}A}\boldsymbol{b})+(P_{\mathrm{Im}A}\boldsymbol{b}-\boldsymbol{b}),$$

上式中，$A\boldsymbol{x}-P_{\mathrm{Im}A}\boldsymbol{b}\in \mathrm{Im}A$ 与 $P_{\mathrm{Im}A}\boldsymbol{b}-\boldsymbol{b}=-(I_m-P_{\mathrm{Im}A})\boldsymbol{b}\in(\mathrm{Im}A)^{\perp}=\mathrm{Ker}A^*$. 因此，

$$\|A\boldsymbol{x}-\boldsymbol{b}\|_2^2=\|A\boldsymbol{x}-P_{\mathrm{Im}A}\boldsymbol{b}\|_2^2+\|P_{\mathrm{Im}A}\boldsymbol{b}-\boldsymbol{b}\|_2^2. \tag{3}$$

由(3)式看出，当且仅当 $A\boldsymbol{x}_0=P_{\mathrm{Im}A}\boldsymbol{b}$ 时，$\boldsymbol{x}_0$ 为 $A\boldsymbol{x}=\boldsymbol{b}$ 的最小二乘解. 另一方面，按 5.1.2 小节(4)式，$X\in A\{1,3\}$等价于 $AX=P_{\mathrm{Im}A}$，因而 $X\boldsymbol{b}$ 为方程 $A\boldsymbol{x}=P_{\mathrm{Im}A}\boldsymbol{b}$ 的解. 于是，$X\boldsymbol{b}$ 为 $A\boldsymbol{x}=\boldsymbol{b}$ 的最小二乘解.

现设 $X=A^{\dagger}$，且 $\boldsymbol{x}_0(\neq A^{\dagger}\boldsymbol{b})$为 $A\boldsymbol{x}=\boldsymbol{b}$ 的最小二乘解. 按照 5.1.2 小节(23)式，$AA^{\dagger}\boldsymbol{b}=P_{\mathrm{Im}A}\boldsymbol{b}$，因而根据 5.1.1 小节推论 8(2)，有 $\boldsymbol{y}\in\mathbb{C}^n$ 使得

$$\boldsymbol{x}_0=A^{\dagger}\boldsymbol{b}+(I_n-A^{\dagger}A)\boldsymbol{y}. \tag{4}$$

上式中，$A^{\dagger}\boldsymbol{b}\in \mathrm{Im}A^{\dagger}=\mathrm{Im}A^*$ 与 $\boldsymbol{0}\neq(I_n-A^{\dagger}A)\boldsymbol{y}\in \mathrm{Ker}A$ 彼此正交，于是，

$$\|\boldsymbol{x}_0\|_2^2=\|A^{\dagger}\boldsymbol{b}\|_2^2+\|(I_n-A^{\dagger}A)\boldsymbol{y}\|_2^2,$$

因此，$\|A^{\dagger}\boldsymbol{b}\|_2<\|\boldsymbol{x}_0\|_2$. □

我们指出，实际上还可以得到如下更好的结论(证明留给读者).

推论 4　设 $A\in \boldsymbol{M}_{m,n}(\mathbb{C})$. 则对任意 $\boldsymbol{b}\in\mathbb{C}^m$，$\boldsymbol{x}=X\boldsymbol{b}$ 为 $A\boldsymbol{x}=\boldsymbol{b}$ 最小二乘解的充分与必要条件是 $X\in A\{1,3\}$. 并且，对任意 $\boldsymbol{b}\in\mathbb{C}^m$，$\boldsymbol{x}=X\boldsymbol{b}$ 为 $A\boldsymbol{x}=\boldsymbol{b}$ 最小欧氏范数的最小二乘解的充分与必要条件是 $X=A^{\dagger}$.

在实际问题(譬如数据曲线拟合)中，矩阵 $A\in \boldsymbol{M}_{m,n}(\mathbb{C})$经常满足 $m\geqslant n$(甚至 $m\gg n$)且 $\mathrm{rank}A=n$(即 A 列满秩). 这时，$A^*A>0$，因而 $\mathrm{Im}A^*=\mathrm{Im}(A^*A)=\mathbb{C}^n$，这相当于：$P_{\mathrm{Im}A^*}=I_n$. 这时 $A^{\dagger}A=P_{\mathrm{Im}A^*}=I_n$. 按(4)式，$\boldsymbol{x}=A^{\dagger}\boldsymbol{b}$ 便是 $A\boldsymbol{x}=\boldsymbol{b}$ 的唯一最小二乘解. 这个结论还可以从另一角度得以证实. 按线性最小二乘法知识，问题(2)有解 $\boldsymbol{x}$ 的充分与必要条件为 $\boldsymbol{x}$ 满足法方程：

$$A^*(A\boldsymbol{x}-\boldsymbol{b})=\boldsymbol{0}. \tag{5}$$

因此，A 列满秩必蕴涵方程组(5)有唯一解 $\boldsymbol{x}=(A^*A)^{-1}A^*\boldsymbol{b}=A^{\dagger}\boldsymbol{b}$ (见 5.1.2 小节定理 15(5)).

现在讨论与线性矩阵方程 $AXB=D$ 有关的最小二乘问题：

$$\|AXB-D\|_F=\min,\quad \|X\|_F=\min \tag{6}$$

的解，式中 $\|\cdot\|_F$ 为 Frobenius 范数(见第二章 2.2.1 小节). 作为定理 3 的推广，对于问题(6)我们有

定理 5　设给定 $A\in \boldsymbol{M}_{m,n}(\mathbb{C})$，$B\in \boldsymbol{M}_{p,q}(\mathbb{C})$ 与 $D\in \boldsymbol{M}_{m,q}(\mathbb{C})$. 则最小二乘问题(6)有唯一解 $X=A^{\dagger}DB^{\dagger}\in \boldsymbol{M}_{n,p}(\mathbb{C})$.

证明　由第四章 4.1.2 小节知道，矩阵方程 $AXB=D$ 等价于线性方程组 $(B^{\mathrm{T}}\otimes A)\mathrm{vec}X=\mathrm{vec}D$，因而问题(6)等价于

$$\|(B^{\mathrm{T}}\otimes A)\mathrm{vec}X-\mathrm{vec}D\|_2=\min,\quad \|\mathrm{vec}X\|_2=\min. \tag{6'}$$

根据定理 3，问题(6′)的唯一解为 $\mathrm{vec}X=(B^{\mathrm{T}}\otimes A)^{\dagger}\mathrm{vec}D=((B^{\dagger})^{\mathrm{T}}\otimes A^{\dagger})\mathrm{vec}D$，此相当于 $X=A^{\dagger}DB^{\dagger}$. □

最后考虑线性矩阵方程

$$AX+YD=E \tag{7}$$

的相容性条件，以及(当相容时)通解的表达式. 式中 $A\in \boldsymbol{M}_{m,n}(\mathbb{C})$，$D\in \boldsymbol{M}_{s,p}(\mathbb{C})$ 与 $E\in \boldsymbol{M}_{m,p}(\mathbb{C})$ 给定，而 $X\in \boldsymbol{M}_{n,p}(\mathbb{C})$ 与 $Y\in \boldsymbol{M}_{m,s}(\mathbb{C})$ 待求. 显然，方程(7)相容等价于存在 X 与 Y 使得

$$AX=E-YD, \tag{8}$$

按 5.1.1 小节定理 7，它又相当于(关于 Y 的)矩阵方程

$$(I_m-AA^{[1]})(E-YD)=O$$

亦即

$$(I_m-AA^{[1]})YD=(I_m-AA^{[1]})E \tag{9}$$

有解 $Y\in \boldsymbol{M}_{m,s}(\mathbb{C})$，这里 $A^{[1]}\in A\{1\}$ 任意给定. 因此，再根据 5.1.1 小节定理 7 可以推得，矩阵方程(9)或(7)相容当且仅当

$$(I_m-AA^{[1]})ED^{[1]}D=(I_m-AA^{[1]})E$$

或等价地

$$(I_m-AA^{[1]})E(I_p-D^{[1]}D)=O, \tag{10}$$

并且当方程(7)相容时，方程(9)的通解表达式为

$$Y=(I_m-AA^{[1]})ED^{[1]}+W-(I_m-AA^{[1]})WDD^{[1]},\qquad \forall W\in \boldsymbol{M}_{m,s}(\mathbb{C}). \tag{11}$$

(这里与(10)式推导中，均用到事实：因为 $I_m-AA^{[1]}$ 为幂等矩阵，所以有 $I_m-AA^{[1]}\in(I_m-AA^{[1]})\{1\}$)将上式代入(8)式右端，并对 X 求解之便得

$$X=A^{[1]}(E-YD)+(I_n-A^{[1]}A)Z,\qquad \forall Z\in \boldsymbol{M}_{n,p}(\mathbb{C}). \tag{12}$$

综上所述，我们有下列关于矩阵方程(7)的结论.

定理 6 设 $A\in \boldsymbol{M}_{m,n}(\mathbb{C})$，$D\in \boldsymbol{M}_{s,p}(\mathbb{C})$与 $E\in \boldsymbol{M}_{m,p}(\mathbb{C})$. 则关于 X 与 Y 的线性矩阵方程(7)相容当且仅当条件(10)成立. 此时，矩阵方程(7)的通解有表达式(11)与(12).

习题 5.1.3

1. 设

$$A=\begin{bmatrix}1 & 1\\ 2 & 0\\ -1 & 3\end{bmatrix},\quad b=\begin{bmatrix}1\\0\\2\end{bmatrix}.$$

求 $A\boldsymbol{x}=\boldsymbol{b}$ 的最小二乘解与具有最小欧氏范数的最小二乘解；求 $A^{\mathrm{T}}\boldsymbol{x}=\boldsymbol{c}$，$\boldsymbol{c}=(1,1)^{\mathrm{T}}$ 的最小二乘解与具有最小欧氏范数的最小二乘解.

2. 设 $A\in \boldsymbol{M}_{m,n}(\mathbb{C})$有非零奇异值 $s_1,s_2,\cdots,s_r$，且 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$ 为 A^*A 在$\mathbb{C}^n$ 中标准正交特征基，$A^*A\boldsymbol{x}_j=s_j^2\boldsymbol{x}_j$，$1\leqslant j\leqslant r$. 试证：$A\boldsymbol{x}=\boldsymbol{b}$ 的具有最小欧氏范数的最小二乘解有表达式

$$\boldsymbol{x}_0 = a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2 + \cdots + a_r\boldsymbol{x}_r,$$

其中，$a_i=(A^*\boldsymbol{b},\boldsymbol{x}_i)/s_i^2$，$i=1,2,\cdots,r$.

3. 试证：矩阵 X 为 A 的 Moore-Penrose 逆当且仅当

$$XAA^* = A^* \text{ 与 } X = BA^*$$

对某矩阵 B 成立.

4. 设 $A_i\in \boldsymbol{M}_{m_i,n}(\mathbb{C})$，$B_i\in \boldsymbol{M}_{p,q_i}(\mathbb{C})$与 $C\in \boldsymbol{M}_{m_i,q_i}(\mathbb{C})$，$i=1,2$. 试证：矩阵方程组 $A_1XB_1=C_1$，$A_2XB_2=C_2$ 有公共解的充分与必要条件为该方程组中每一个矩阵方程相容，且对某对 $A_i^{[1]}\in A_i\{1\}$与 $B_i^{[1]}\in B_i\{1\}$，$i=1,2$ 与满足

$$\mathrm{Im}A = \mathrm{Im}A_1 \cap \mathrm{Im}A_2,\ \mathscr{C}(B) = \mathscr{C}(B_1) \cap \mathscr{C}(B_2)$$

(这里，$\mathscr{C}(B)$为 B 的行空间)的 A 与 B 有

$$AA_1^{[1]}C_1B_1^{[1]}B = AA_2^{[1]}C_2B_2^{[1]}B.$$

*5. 给出线性矩阵方程

$$AXB + CYD = E$$

的相容性条件以及通解的表达式，式中，$A\in \boldsymbol{M}_{m,n}(\mathbb{C})$，$B\in \boldsymbol{M}_{p,q}(\mathbb{C})$，$C\in \boldsymbol{M}_{m,r}(\mathbb{C})$，$D\in \boldsymbol{M}_{s,q}(\mathbb{C})$与 $E\in \boldsymbol{M}_{m,q}(\mathbb{C})$给定，而 $X\in \boldsymbol{M}_{n,p}(\mathbb{C})$与 $Y\in \boldsymbol{M}_{r,s}(\mathbb{C})$待求(试与定理 6 比较你的结果).

6. 考虑两个相容的线性代数方程组

$$A_1\boldsymbol{x} = \boldsymbol{b}_1,A_2\boldsymbol{x} = \boldsymbol{b}_2.$$

试证：它们有公共解 $\boldsymbol{x}$ 当且仅当下列方程组相容：

$$A_2P_{\mathrm{Ker}A_1}\boldsymbol{y} = \boldsymbol{b}_2 - A_2A_1^{[1,4]}\boldsymbol{b}_1.$$

此时，它们的公共通解有表达式

$$\boldsymbol{x} = A_1^{[1,4]}\boldsymbol{b}_1 + P_{\mathrm{Ker}A_1}(A_2P_{\mathrm{Ker}A_1})^{[1]}(\boldsymbol{b}_2 - A_2A_1^{[1,4]}\boldsymbol{b}_1) + \mathrm{Ker}A_1 \cap \mathrm{Ker}A_2,$$

或

$$\boldsymbol{x} = A_2^{[1,4]}\boldsymbol{b}_2 + P_{\mathrm{Ker}A_2}(A_1P_{\mathrm{Ker}A_2})^{[1]}(\boldsymbol{b}_1 - A_1A_2^{[1,4]}\boldsymbol{b}_2) + \mathrm{Ker}A_2 \cap \mathrm{Ker}A_1,$$

其中，$P_{\mathrm{Ker}A_i}$表示在 $\mathrm{Ker}A_i$ 上的正交投影，$i=1,2$.

7. 设 $A\in \boldsymbol{M}_{m,n}(\mathbb{C})$，$\boldsymbol{b}\in \mathbb{C}^m$，且 $\mathscr{S}$为$\mathbb{C}^n$ 的子空间. 试证：受约束线性系统

$$Ax = b, \qquad x \in \mathcal{S}$$

相容当且仅当 $AP_{\mathcal{S}}z=b$ 相容. 此时,该系统的通解有表达式

$$x = Xb + (I_n - XA)y, \qquad \forall\, y \in \mathcal{S},$$

其中,$X=P_{\mathcal{S}}(AP_{\mathcal{S}})^{(1)}$,$P_{\mathcal{S}}$ 表示在 $\mathcal{S}$上的正交投影.

8. 设 $A\in M_n(\mathbb{C})$,$b\in\mathbb{C}^n$,且 $\mathcal{S}$为$\mathbb{C}^n$ 的子空间. 试证:受约束系统

$$Ax + y = b, \qquad x \in \mathcal{S}, y \in \mathcal{S}^{\perp} \tag{$*$}$$

相容当且仅当下列(关于 z 的)方程组相容:

$$(AP_{\mathcal{S}} + P_{\mathcal{S}^{\perp}})z = b. \tag{$**$}$$

并且,x 与 y 为系统($*$)的解当且仅当

$$x = P_{\mathcal{S}}z, \qquad y = P_{\mathcal{S}^{\perp}}z = b - AP_{\mathcal{S}}z,$$

这里 z 为方程组($**$)的解. 假如 $AP_{\mathcal{S}}+P_{\mathcal{S}^{\perp}}\in M_n(\mathbb{C})$非奇异,那么,对所有 $b\in\mathbb{C}^n$,系统($*$)有唯一解

$$x = P_{\mathcal{S}}(AP_{\mathcal{S}} + P_{\mathcal{S}^{\perp}})^{-1}b, y = b - Ax.$$

9. 试证:对任意 $A\in M_n(\mathbb{C})$与$\mathbb{C}^n$ 的子空间 $\mathcal{S}$,下列诸命题彼此等价:

(1) $AP_{\mathcal{S}}+P_{\mathcal{S}^{\perp}}$ 非奇异.

(2) $\mathbb{C}^n=A\mathcal{S}\dot{+}\mathcal{S}^{\perp}$,即 $A\mathcal{S}=\{Ax:x\in\mathcal{S}\}$与 $\mathcal{S}^{\perp}$ 为$\mathbb{C}^n$ 的互补子空间.

(3) $\mathbb{C}^n=P_{\mathcal{S}}\mathrm{Im}A\dot{+}\mathcal{S}^{\perp}$.

(4) $\mathbb{C}^n=P_{\mathcal{S}}A\mathcal{S}\dot{+}\mathcal{S}^{\perp}$.

(5) $\mathrm{rank}(P_{\mathcal{S}}AP_{\mathcal{S}})=\dim\mathcal{S}$.

5.2　方阵的谱广义逆

本节讨论具有一般非奇异矩阵逆的某些谱性质的所谓**谱广义逆**,它们仅适用于方阵的情形,因为只有方阵才有特征值与特征向量.

5.2.1　Drazin 逆

除了在本章前言部分介绍过的四个 Penrose 方程

$$AXA = A, \tag{1}$$

$$XAX = X, \tag{2}$$

$$AX = (AX)^*, \tag{3}$$

$$XA = (XA)^* \tag{4}$$

以外,我们还考虑只适用于方阵的如下方程:

$$A^kXA = A^k, \tag{1^k}$$

$$AX = XA, \tag{5}$$

$$A^kX = XA^k, \tag{5^k}$$

$$A^k = XA^{k+1}, \tag{6^k}$$

这里,k 为某个给定的非负整数. 如 5.1.1 小节所述,对这八个方程也可以定义方阵 A 的 λ-逆,其中 λ 可以是集合 $\{1,2,3,4,1^k,5,5^k,6^k\}$ 的任意不空子集. 譬如,方阵 A 的 $\{1^k,2,5\}$-逆意味着,它满足方程 (1^k),(2)与(5).

定义 1　$A\in \boldsymbol{M}_n(\mathbb{C})$的指标,index$(A)$,指的是 A 的零特征值的指标,即满足 $\mathrm{rank}A^{k+1}=\mathrm{rank}A^k$ 的最小非负整数 k,或等价地,A 的最小多项式 $m_A(\lambda)$中因子 λ^k 的指数 k.

按照定义 1,非奇异矩阵 A 的指标 index$(A)=0$,而奇异矩阵的指标总是$\geqslant 1$. 例如,对二阶矩阵

$$A=\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix},$$

index$(A)=2$,因为 $m_A(\lambda)=\lambda^2$ 或 $\mathrm{rank}A=1$,$\mathrm{rank}A^2=\mathrm{rank}A^3=0$.

假如 $A\in \boldsymbol{M}_n(\mathbb{C})$有 index$(A)=k$,那么由第一章 1.1.2 小节(13)式知道,此时且仅在此时,$\mathrm{Im}A^l=\mathrm{Im}A^k$ 与 $\mathrm{Ker}A^l=\mathrm{Ker}A^k$, $\forall\, l\geqslant k$,因而 $\mathrm{rank}A^l=\mathrm{rank}A^k$ 与 $\mathrm{nul}A^l=\mathrm{nul}A^k$, $\forall\, l\geqslant k$. 于是,index$(A)=k$ 蕴涵对任意 $l\geqslant k$,矩阵方程

$$A^{l+1}X=A^l \tag{7}$$

有解 $X\in \boldsymbol{M}_n(\mathbb{C})$. 反之,使得方程(7)有解 X 的最小非负整数 $l=k$ 就是 index(A). 事实上,(7)式蕴涵 $\mathrm{rank}A^{l+1}=\mathrm{rank}A^l$, $\forall\, l\geqslant k$,再按 $\mathrm{Im}A^{l+1}\subset \mathrm{Im}A^l$, $\forall\, l\geqslant k$ 立即推出 $\mathrm{Im}A^{l+1}=\mathrm{Im}A^l$, $\forall\, l\geqslant k$,另一方面,由 k 的最小性看出,$\mathrm{Im}A^{l+1}\neq \mathrm{Im}A^l$, $\forall\, l<k$. 因此,index$(A)=k$.

综上所述,我们有关于 A 的指标 index(A)的如下结果.

定理 2　设 $A\in \boldsymbol{M}_n(\mathbb{C})$. 则 A 的指标 index(A)等于使(7)式成立的最小非负整数 l.

有了矩阵指标的概念,可以引入 Drazin 逆的定义.

定义 3　若 $A\in \boldsymbol{M}_n(\mathbb{C})$有 index$(A)=k$,则满足方程$(1^k)$,(2),(5)的矩阵 $X\in \boldsymbol{M}_n(\mathbb{C})$称为 A 的 **Drazin 逆**.

容易看出方程(1^k),(2)与(5)总体上等价于(2),(5)与(6^k). 因此,定义 3 中方程(1^k)有时也用方程(6^k)代替.

下面基本定理指出,$A\in \boldsymbol{M}_n(\mathbb{C})$的 Drazin 逆是存在唯一的,且它可表示为 A 的多项式.

定理 4　设 $A\in \boldsymbol{M}_n(\mathbb{C})$有 index$(A)=k$,且 A 的最小多项式为 $m(\lambda)=c\lambda^k(1-\lambda q(\lambda))$,这里,$c\neq 0$ 为常数,$q(\lambda)$为多项式. 则 A 有唯一的 Drazin 逆 X,它可以表示为关于 A 的多项式:

$$X=A^k(q(A))^{k+1} \tag{8}$$

证明　唯一性:设 X 与 Y 为 A 的两个 Drazin 逆. 令 $E=AX=XA$ 与 $F=AY=YA$,则 E 与 F 皆为幂等矩阵. 此时,

$$E = AX = A^kX^k = AYA^kX^k = FAX = FE,$$
$$F = AY = Y^kA^k = Y^kA^kXA = YAE = FE.$$

因此,$E=F$,从而 $X=EX=FX=YE=YF=Y$,因为 $FX=YAX=YE$.

存在性:$m(A)=O$ 蕴涵 $A^{k+1}q(A)=A^k$. 现令 $X=A^k(q(A))^{k+1}$. 它显然满足方程(5),并且,

$$\begin{aligned} XAX &= A^{2k+1}(q(A))^{2k+2} = A^k(A^{k+1}q(A))(q(A))^{2k+1} \\ &= A^{2k}(q(A))^{2k+1} = \cdots = A^k(q(A))^{k+1} = X, \\ XA^{k+1} &= A^{k+1}X = A^{2k+1}(q(A))^{k+1} = A^k(A^{k+1}q(A))q(A)^k \\ &= A^{2k}(q(A))^k = \cdots = A^k. \end{aligned}$$

因此按定义,X 为 A 的 Drazin 逆. □

今后,将 $A\in \boldsymbol{M}_n(\mathbb{C})$的唯一的 Drazin 逆记为 A^{D}.

按定义,$AA^{\mathrm{D}}=A^{\mathrm{D}}A$ 为幂等矩阵,因而它们是沿着 $\mathrm{Ker}(A^{\mathrm{D}}A)$在 $\mathrm{Im}(A^{\mathrm{D}}A)$上的投影. 但由 $A^{\mathrm{D}}=A^k(q(A))^{k+1}$(见公式(8))与 $A^{\mathrm{D}}A^{k+1}=A^k$ 以及 5.1.2 小节定理 1(2)与(5)可立即推出,若 $k=\mathrm{index}(A)$,则

$$\mathrm{Im}A^{\mathrm{D}} = \mathrm{Im}(A^{\mathrm{D}}A) = \mathrm{Im}A^k, \quad \mathrm{Ker}A^{\mathrm{D}} = \mathrm{Ker}(A^{\mathrm{D}}A) = \mathrm{Ker}A^k. \tag{9}$$

因此,按 5.1.2 小节定理 1(3),

$$\mathbb{C}^n = \mathrm{Im}A^k \dotplus \mathrm{Ker}A^k, \qquad k = \mathrm{index}(A), \tag{10}$$

这表明,$AA^{\mathrm{D}}=A^{\mathrm{D}}A$ 便是沿着 $\mathrm{Ker}A^k$ 在 $\mathrm{Im}A^k$ 上的投影. 特殊地,当 $A=A^*$ 时,$AA^{\mathrm{D}}=A^{\mathrm{D}}A$ 是 $\mathrm{Im}A^k$ 上的正交投影.

我们指出,当 $k=\mathrm{index}(A)$时,用类似于定理 4 的证明可得出,$Y=A^l(q(A))^{l+1}$($\forall\, l\geqslant k$)也满足方程(2),(5)与(6^k),因而由 A 的 Drazin 逆的唯一性便知

$$A^{\mathrm{D}} = A^l(q(A))^{l+1}, \qquad \forall\, l \geqslant k, \tag{11}$$

这里,$q(\lambda)$的涵义同定理 4(见习题 3).

A^{D} 的其余性质罗列在下面命题中,其证明留给读者(见习题 4).

命题 5　设 $A\in \boldsymbol{M}_n(\mathbb{C})$有 $\mathrm{index}(A)=k$. 则

(1) $(A^*)^{\mathrm{D}}=(A^{\mathrm{D}})^*$,$(A^{\mathrm{T}})^{\mathrm{D}}=(A^{\mathrm{D}})^{\mathrm{T}}$.

(2) $(A^l)^{\mathrm{D}}=(A^{\mathrm{D}})^l$,$\forall\, l=1,2,\cdots$.

(3) $(A^{\mathrm{D}})^{\mathrm{D}}=A$ 当且仅当 $\mathrm{index}(A)=k\leqslant 1$.

(4) $((A^{\mathrm{D}})^{\mathrm{D}})^{\mathrm{D}}=A^{\mathrm{D}}$,且 $\mathrm{index}(A^{\mathrm{D}})\leqslant 1$.

(5) 若 A 幂零,则 $A^{\mathrm{D}}=O$.

(6) 若 $l>m>0$,则 $A^m(A^{\mathrm{D}})^l=(A^{\mathrm{D}})^{l-m}$.

(7) 若 $m>0$ 与 $l-m\geqslant k$,则 $A^l(A^{\mathrm{D}})^m=A^{l-m}$.

(8) $A^{\mathrm{D}}=A^l(A^{2l+1})^{[1]}A^l$,$\forall\, l\geqslant k$,其中$(A^{2l+1})^{[1]}$为 A^{2l+1}的任意$\{1\}$-逆.

(9) 设 $Q\in \boldsymbol{M}_n(\mathbb{C})$非奇异,且 $A=QBQ^{-1}$,则 $A^{\mathrm{D}}=QB^{\mathrm{D}}Q^{-1}$.

例 6　设 $J\in \boldsymbol{M}_7(\mathbb{C})$ 有形式

$$J=\left[\begin{array}{cc:ccc:cc} 0 & 1 & & & & & \\ 0 & 0 & & 0 & & \multicolumn{2}{c}{0} \\ \hdashline & & 0 & 1 & 0 & & \\ \multicolumn{2}{c:}{0} & 0 & 0 & 1 & \multicolumn{2}{c}{0} \\ & & 0 & 0 & 0 & & \\ \hdashline \multicolumn{2}{c:}{0} & & 0 & & 1 & 1 \\ & & & & & 0 & 1 \end{array}\right] \tag{12}$$

求 J^{D}.

显然 J 的最小多项式为 $m(\lambda)=\lambda^3(\lambda-1)^2=\lambda^3(1-\lambda(2-\lambda))$，于是，定理 4 中 $q(\lambda)=2-\lambda$，$\operatorname{index}(J)=3$. 因此，$J^{\mathrm{D}}=J^3(q(J))^4=J^3(2I_7-J)^4$. 经过简单计算可得

$$J^{\mathrm{D}}=\operatorname{diag}\left[O_2,O_3\begin{bmatrix}1 & -1\\ 0 & 1\end{bmatrix}\right], \tag{13}$$

这里，O_2 与 O_3 分别为二阶与三阶零矩阵. 实际上，由直接验证或应用命题 5(5)与习题 1 也可得(13)式，因为我们有

$$\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}^{\mathrm{D}}=\begin{bmatrix}0 & 0\\ 0 & 0\end{bmatrix},\ \begin{bmatrix}0 & 1 & 0\\ 0 & 0 & 1\\ 0 & 0 & 0\end{bmatrix}^{\mathrm{D}}=\begin{bmatrix}0 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix},$$

$$\begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}^{\mathrm{D}}=\begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}^{-1}=\begin{bmatrix}1 & -1\\ 0 & 1\end{bmatrix}.$$ □

习题 5.2.1

1. 设 $A\in \boldsymbol{M}_n(\mathbb{C})$ 有分块对角矩阵形式 $A=\operatorname{diag}[A_1,A_2,\cdots,A_p]$. 试证：

$$A^{\mathrm{D}}=\operatorname{diag}[A_1^{\mathrm{D}},A_2^{\mathrm{D}},\cdots,A_p^{\mathrm{D}}].$$

2. 试证：若 $A\in \boldsymbol{M}_n(\mathbb{C})$ 有 $\operatorname{index}(A)=k$，则 A 的 Drazin 逆 A^{D} 为由下式给出的唯一矩阵：

$$A^{\mathrm{D}}\boldsymbol{x}=\begin{cases}\boldsymbol{y}, \text{如果 } A\boldsymbol{y}=\boldsymbol{x},\boldsymbol{x}\in \operatorname{Im}A^k,\\ \boldsymbol{0}, \text{如果 } A^k\boldsymbol{x}=\boldsymbol{0}.\end{cases}$$

3. 设 $A\in \boldsymbol{M}_n(\mathbb{C})$ 有 $\operatorname{index}(A)=k$. 试证：对任意 $l\geqslant k$，A 的 $\{1^l,2,5\}$-逆是唯一存在的，且等于 A^{D}.

4. 证明命题 5.

5. 设 $B\in \boldsymbol{M}_{m,n}(\mathbb{C})$ 与 $W\in \boldsymbol{M}_{n,m}(\mathbb{C})$. 试证：

(1) $(BW)^{\mathrm{D}}=B((WB)^{\mathrm{D}})^2W$.

(2) $W((BW)^{\mathrm{D}})^p=((WB)^{\mathrm{D}})^pW$，$B((WB)^{\mathrm{D}})^p=((BW)^{\mathrm{D}})^pB$，$\forall p=0,1,2,\cdots$.

(3) 对任意给定正整数 p，存在唯一 $X\in \boldsymbol{M}_{m,n}(\mathbb{C})$ 满足

$$(BW)^{\mathrm{D}}XW=((BW)^{\mathrm{D}})^p,$$

$$BWX = XWB,$$
$$BW(BW)^{D}X = X.$$

同时,存在唯一 $X \in \boldsymbol{M}_{m,n}(\mathbb{C})$满足

$$XW = BW((BW)^{D})^{p},$$
$$WX = WB((WB)^{D})^{p},$$
$$XW(BW)^{p-1}X = X.$$

并且,满足上述六个方程的唯一 $X \in \boldsymbol{M}_{m,n}(\mathbb{C})$为

$$X = B((WB)^{D})^{p}.$$

6. 设 B 与 W 同题 5. 试证:$X=B((WB)^{D})^{2}$ 为下列三个方程的唯一公共解:

$$(BW)^{k} = (BW)^{k+1}XW, k \text{ 为某正整数},$$
$$X = XWBWX,$$
$$BWX = XWB.$$

7. 设 B 与 W 同题 5. 称题 6 中 $X=B((WB)^{D})^{2} \in \boldsymbol{M}_{m,n}(\mathbb{C})$为 B 的 **W-加权 Drazin 逆**,记为 B_{W}^{D}.(显然,当 $B \in \boldsymbol{M}_{n}(\mathbb{C})$与 $W=I_{n}$ 时,B_{W}^{D} 即为 B 的 Drazin 逆 B^{D})试证:$X=B_{W}^{D}$,$W \in \boldsymbol{M}_{n,m}(\mathbb{C})$,等价于 X 有表达式

$$X = BYBYB,$$

式中,$Y \in \boldsymbol{M}_{n,m}(\mathbb{C})$是满足 $\mathrm{index}(BY)=\mathrm{index}(YB) \leqslant 1$ 的某个矩阵.

8. 设 $A \in \boldsymbol{M}_{p}(\mathbb{C})$,$B \in \boldsymbol{M}_{p,q}(\mathbb{C})$,$C=\boldsymbol{M}_{q,p}(\mathbb{C})$与 $E \in \boldsymbol{M}_{q}(\mathbb{C})$,并且,$\mathrm{index}(A)=k$. 证明:如果 $\mathrm{Im}B \subset \mathrm{Im}A^{k}$ 与 $\mathrm{Im}C^{*} \subset \mathrm{Im}((A^{k})^{*})$,则

$$\begin{pmatrix} A & B \\ C & E \end{pmatrix} = \begin{pmatrix} I_{p} & O \\ CA^{D} & I_{q} \end{pmatrix} \begin{pmatrix} A & O \\ O & E-CA^{D}B \end{pmatrix} \begin{pmatrix} I_{p} & A^{D}B \\ O & I_{q} \end{pmatrix}.$$

5.2.2 群逆与广义左(右)逆

先考虑 $A \in \boldsymbol{M}_{n}(\mathbb{C})$的群逆.

定义 1 矩阵 $X \in \boldsymbol{M}_{n}(\mathbb{C})$称为 $A \in \boldsymbol{M}_{n}(\mathbb{C})$的**群逆**(group inverse),假如它满足 5.2.1 小节中方程(1),(2)与(5).

显然,由定义 1 看出,X 为 A 的群逆等价于 A 为 X 的群逆. 若 A 的群逆 X 存在,则 $X \in A\{1,2\}$,因而按条件 $AX=XA$ 与 5.1.2 小节定理 2,

$$\mathrm{Im}X = \mathrm{Im}A = \mathrm{Im}(AX), \qquad \mathrm{Ker}X = \mathrm{Ker}A = \mathrm{Ker}(AX), \tag{1}$$

因而

$$\mathbb{C}^{n} = \mathrm{Im}A \dotplus \mathrm{Ker}A. \tag{2}$$

但是,不同于 $A \in \boldsymbol{M}_{n}(\mathbb{C})$的 Drazin 逆 A^{D},A 的群逆可能不存在. 例如 5.2.1 小节(12)式中矩阵 J 就没有群逆,因为如存在 $X \in \boldsymbol{M}_{n}(\mathbb{C})$使得 $JXJ=J$,$XJX=X$,$JX=XJ$,则有 $J^{3}XJ=J^{3}$,$XJX=X$,$JX=XJ$,因而 $X=J^{D}$. 但从 5.2.1 小节(13)式知道,$JJ^{D}J \neq J$. 因而 J^{D} 不是 J 的群逆,这表明 J 没有群逆.

假如 A 的指标 $\mathrm{index}(A) \leqslant 1$,按定义,$A$ 的群逆此时就是 A 的 Drazin 逆 A^{D},因而它一定唯一地存在. 下面关于 A 的群逆的基本定理指出,$\mathrm{index}(A) \leqslant 1$ 也是 A

有群逆的必要条件. 这个定理同时还给出 A 有群逆的另外两个等价条件与群逆表达式.

定理 2　设 $A\in \boldsymbol{M}_n(\mathbb{C})$ 有秩 $r>0$,且有满秩分解 $A=FG$. 则下列诸命题彼此等价:

(1) A 有群逆.

(2) $\mathbb{C}^n=\mathrm{Im}A\dot{+}\mathrm{Ker}A$,即 $\mathrm{Im}A$ 与 $\mathrm{Ker}A$ 在 $\mathbb{C}^n$ 中互补.

(3) $\mathrm{index}(A)\leqslant 1$.

(4) $GF\in \boldsymbol{M}_r(\mathbb{C})$ 非奇异.

并且,当 A 有群逆 X 时,X 唯一且可表示为

$$X = F(GF)^{-2}G = A(q(A))^2, \tag{3}$$

式中,$q(\lambda)$ 的涵义同 5.2.1 小节定理 4.

证明　当 $\mathrm{index}(A)\leqslant 1$ 时,群逆存在唯一性,(1)蕴涵(2)诸条前面已证.

(2)⇒(3):由于 $\mathrm{rank}A+\mathrm{nul}A=n$,故(2)等价于

$$\mathrm{Im}A \cap \mathrm{Ker}A = \{\mathbf{0}\}.$$

若 A 非奇异,则 $\mathrm{index}(A)=0$;若 A 奇异且 $\mathrm{index}(A)>1$,则有 $\mathrm{rank}A^2<\mathrm{rank}A$,因而 $\mathrm{Ker}A\subsetneqq \mathrm{Ker}A^2$. 这表明有 $\boldsymbol{x}\neq\mathbf{0}$,使得 $A^2\boldsymbol{x}=\mathbf{0}$ 但 $A\boldsymbol{x}\neq\mathbf{0}$,此等价于有非零向量 $A\boldsymbol{x}\in \mathrm{Im}A\cap \mathrm{Ker}A$. 这个矛盾推出 $\mathrm{index}(A)=1$.

(3)⇒(4):若 $\mathrm{index}(A)\leqslant 1$,则从 $A^2=FGFG$ 我们推出,

$$r = \mathrm{rank}A = \mathrm{rank}A^2 \leqslant \mathrm{rank}(GF) \leqslant \mathrm{rank}F = r,$$

因而 $GF\in \boldsymbol{M}_r(\mathbb{C})$ 非奇异.

(4)⇒(1):设 $\mathrm{rank}(GF)=r$,则(3)式中间项有意义. 直接验证表明,由(3)式中第一个等式确定的 X 满足 5.2.1 小节中方程(1),(2)与(5). 因此,它是 A 的群逆. 此外,$X=A(q(A))^2$ 为 5.2.1 小节定理 4 中公式 $A^{\mathrm{D}}=A^k(q(A))^{k+1}$ 当 $k\leqslant 1$ 时的特殊情形. 因此,它也为 A 的群逆. 由 A 的群逆的唯一性便得(3)式. □

通常,用记号 $A^{\#}$ 表示 $A\in \boldsymbol{M}_n(\mathbb{C})$ 的群逆. 当 $\mathrm{rank}A=0$ 时,显然 $A^{\#}=O$,而当 $\mathrm{rank}A=n$ 时,$A^{\#}=A^{-1}$. 若 A 奇异但 $\mathrm{index}(A)=1$,则 $A^{\#}=A^{\mathrm{D}}$. 因此,群逆恰是指标不超过 1 的矩阵的 Drazin 逆的特殊情形.

对照 5.2.1 小节命题 5,群逆还有一些重要性质,罗列如下(证明留给读者).

命题 3　设 $A\in \boldsymbol{M}_n(\mathbb{C})$ 有 $\mathrm{index}(A)\leqslant 1$. 则

(1) $(A^{\#})^{\#}=A$.

(2) $(A^*)^{\#}=(A^{\#})^*$, $(A^{\mathrm{T}})^{\#}=(A^{\#})^{\mathrm{T}}$.

(3) $(A^l)^{\#}=(A^{\#})^l$, $\forall\, l=1,2,\cdots$.

(4) 令 $A^{-j}=(A^{\#})^j$, $j=1,2,\cdots$; $A^0=AA^{\#}=A^{\#}A$. 则对所有整数 l 与 m, $A^lA^m=A^{l+m}$.

(5) $A^{\#}=A(A^3)^{[1]}A$, 这里 $(A^3)^{[1]}$ 为 A^3 的任意{1}-逆.

我们指出，上述命题(4)表明，在那里的约定下，A 的"幂"(正的，负的，零次幂)关于矩阵乘法组成一个可换群. 这正是群逆称谓的源由.

下面定理说明，在一定条件下，$A^{\#}$ 可以等于 $A^{\dagger}$.

定理 4　设 $A\in \boldsymbol{M}_n(\mathbb{C})$. 则下面诸命题彼此等价：

(1) A 为**值域 Hermite 矩阵**，即 A 满足

$$\mathrm{Im}A^* = \mathrm{Im}A. \tag{4}$$

(2) $A^{\#}$ 存在且 $A^{\#}=A^{\dagger}$.

(3) $AA^{\dagger}=A^{\dagger}A$.

(4) 对某 $Y\in \boldsymbol{M}_n(\mathbb{C})$，$A^*=YA$.

证明　(1)⇒(2)：由于 $\mathrm{Im}A^*=(\mathrm{Ker}A)^{\perp}$，故(1)等价于 $\mathrm{Ker}A^*=\mathrm{Ker}A$，亦即 $\mathrm{Im}A\perp\mathrm{Ker}A$. 按定理 2，$A^{\#}$ 存在. 这时按 5.1.2 小节推论 6，$X=A^{[1,2]}_{\mathrm{Im}A,\mathrm{Ker}A}$ 是 $A\{1,2\}$ 中满足条件 $\mathrm{Im}X=\mathrm{Im}A$ 与 $\mathrm{Ker}X=\mathrm{Ker}A$ 的唯一元素，对此 X，$AX=P_{\mathrm{Im}A,\mathrm{Ker}X}=P_{\mathrm{Im}X,\mathrm{Ker}A}=XA$(见 5.1.2 小节(1)式). 因此，按定义此 X 等于 $A^{\#}$. 并且 $A^{\#}=A^{[1,2]}_{\mathrm{Im}A,\mathrm{Ker}A}=A^{[1,2]}_{\mathrm{Im}A^*\,\mathrm{Ker}A^*}=A^{\dagger}$. (2)⇒(3)：如果 $A^{\#}=A^{\dagger}$，则 $AA^{\dagger}=AA^{\#}=A^{\#}A=A^{\dagger}A$.

(1)⇔(3)：这是显然的，因为(4)式等价于 $P_{\mathrm{Im}A^*}=P_{\mathrm{Im}A}$，而后者又等价于 $A^{\dagger}A=AA^{\dagger}$.

(1)⇔(4)：设(1)成立. 则有矩阵 Z 使得 $A=A^*Z$，因而 $A^*=Z^*A$. 取 $Y=Z^*$ 便导出(4). 反之，(4)蕴涵 $A=A^*Y^*$ 与 $\mathrm{rank}A^*\leqslant\mathrm{rank}A$，因而 $\mathrm{Im}A\subset\mathrm{Im}A^*$ 与 $\mathrm{rank}A^*\leqslant\mathrm{rank}A$，由此即得(4)式. □

从定理 4 与定理 2 看出，值域 Hermite 矩阵的指标不超过 1. 显然，所有非奇异矩阵 $A\in \boldsymbol{M}_n(\mathbb{C})$ 为值域 Hermite 矩阵，因为 $\mathrm{Im}A^*=\mathrm{Im}A=\mathbb{C}^n$. 此外，任意正规矩阵(因而任意 Hermite 矩阵与酉矩阵)也是值域 Hermite 矩阵，这是因为 $\mathrm{Im}(AA^*)=\mathrm{Im}A$ 与 A 正规蕴涵 $\mathrm{Im}A=\mathrm{Im}A^*$. 但是，指标不超过 1 的矩阵 $A\in \boldsymbol{M}_n(\mathbb{C})$ 不一定为值域 Hermite 矩阵，例如，

$$A=\begin{bmatrix}1 & 0\\ 1 & 0\end{bmatrix}.$$

我们指出，群逆具有可以与非奇异矩阵的谱性质(即与特征值、特征向量以及广义特征向量有关的性质)相比较的一些性质. 譬如，当 $A\in \boldsymbol{M}_n(\mathbb{C})$ 为简单矩阵时，存在有非奇异矩阵 P 使得 $AP=PJ$，这里 $J=\mathrm{diag}[\lambda_1,\lambda_2,\cdots,\lambda_n]$ 为 A 的 Jordan 正规形式. 此时，$\mathrm{index}(J)\leqslant 1$，$J^{\#}=J^{\dagger}$. 可以直接验证，$A^{\#}=PJ^{\#}P^{-1}=PJ^{\dagger}P^{-1}=P\mathrm{diag}[\lambda_1^{\dagger},\lambda_2^{\dagger},\cdots,\lambda_n^{\dagger}]P^{-1}$，这里，$\lambda_j^{\dagger}$ 涵义同 5.1.1 小节(6)式. 这说明 $A^{\#}$ 有特征值 $\lambda_1^{\dagger},\lambda_2^{\dagger},\cdots,\lambda_n^{\dagger}$，其对应的特征向量与 A 的特征值 $\lambda_1,\lambda_2,\cdots,\lambda_n$ 对应的特征向量相同，即为矩阵 P 的 n 个列向量. 对于一般矩阵 $A\in \boldsymbol{M}_n(\mathbb{C})$，当 $\mathrm{index}(A)\leqslant 1$ 时可以证明：对任意 $\lambda\in\mathbb{C}$ 与 $\boldsymbol{x}\in\mathbb{C}^n$，以及正整数 p，

$$(A-\lambda I)^p\boldsymbol{x}=\boldsymbol{0},(A-\lambda I)^{p-1}\boldsymbol{x}\neq\boldsymbol{0}\Leftrightarrow(A^{\#}-\lambda^{\dagger}I)^p\boldsymbol{x}=\boldsymbol{0},(A^{\#}-\lambda^{\dagger}I)^{p-1}\boldsymbol{x}\neq\boldsymbol{0}.$$

换句话说,x 为 A 的相应于特征值 λ 的 p 级广义特征向量当且仅当 x 为 $A^{\#}$ 的相应于特征值 $\lambda^{\dagger}$ 的 p 级广义特征向量(见习题 2).

接着介绍**广义左(右)逆**的概念与特性.

在第一章 1.1.2 小节中,我们介绍了矩阵 $A\in\boldsymbol{M}_{m,n}(\mathbb{C})$ 的左逆与右逆概念与性质. 回忆一下,Y 为 A 的左(右)逆,假如 $YA=I_n(AY=I_m)$. 对于方阵 $A\in\boldsymbol{M}_n(\mathbb{C})$ 来说,A 的左逆等于 A 的右逆,且二者都等于 A 的逆 A^{-1}. 对于一般矩阵 $A\in\boldsymbol{M}_{m,n}(\mathbb{C})$,当且仅当 A 列满秩(行满秩)时,其左(右)逆存在(第一章 1.1.2 小节定理 1). 这时,A 的左逆或右逆均为 A 的{1,2}-逆. 下面引入方阵的广义左逆与广义右逆的定义.

定义 5 设 $A\in\boldsymbol{M}_n(\mathbb{C})$. 若 $Y\in\boldsymbol{M}_n(\mathbb{C})$满足

$$YA\boldsymbol{x}=\boldsymbol{x},\qquad \forall\,\boldsymbol{x}\in\bigcap_{m=0}^{\infty}\mathrm{Im}A^m,\tag{5}$$

则称 Y 为 A 的一个**广义左逆**. 若 $Z\in\boldsymbol{M}_n(\mathbb{C})$满足

$$\boldsymbol{x}^*AZ=\boldsymbol{x}^*,\qquad \forall\,\boldsymbol{x}\in\bigcap_{m=0}^{\infty}\mathrm{Im}(A^*)^m,\tag{6}$$

则称 Z 为 A 的一个**广义右逆**.

显然,Z 为 A 的一个广义右逆当且仅当 Z^* 为 A^* 的一个广义左逆. 当 $B\in\boldsymbol{M}_n(\mathbb{C})$有指标 $\mathrm{index}(B)=k$ 时,定义 5 中

$$\bigcap_{m=0}^{\infty}\mathrm{Im}B^m=\mathrm{Im}B^k.$$

下面定理给出广义左逆的基本特征. 对于广义右逆,显然也有相应的结果(见习题 3).

定理 6 设 $A,Y\in\boldsymbol{M}_n(\mathbb{C})$. 则下列诸命题彼此等价:

(1) Y 为 A 的一个广义左逆.

(2) $YA^{k+1}=A^k$,这里 $k=\mathrm{index}(A)$.

(3) $YA^{l+1}=A^l$, $\forall\,l\geqslant k=\mathrm{index}(A)$.

(4) $YA^{l+1}=A^l$ 对某个 $l\geqslant0$ 成立.

证明 (1)⇒(2):设 Y 为 A 的一个广义左逆,且 A 的指标为 k,则按定义 5,对任意 $\boldsymbol{x}\in\mathbb{C}^n$, $A^k\boldsymbol{x}\in\mathrm{Im}A^k=\bigcap_{m=0}^{\infty}\mathrm{Im}A^m$,因而 $YA^{k+1}\boldsymbol{x}=YA(A^k\boldsymbol{x})=A^k\boldsymbol{x}$,由此即得(2).

(2)⇒(3)⇒(4)是显然的.

(4)⇒(1):假定对某个 $l\geqslant0$, $YA^{l+1}=A^l$. 由于 $\mathrm{rank}A^l=\mathrm{rank}(YA^{l+1})\leqslant\mathrm{rank}A^{l+1}\leqslant\mathrm{rank}A^l$,故有

$$\mathrm{rank}A^s=\mathrm{rank}A^l,\qquad \mathrm{Im}A^s=\mathrm{Im}A^l,\ \forall\,s\geqslant l.$$

现任取 $\boldsymbol{x}\in\bigcap_{m=0}^{\infty}\mathrm{Im}A^m=\mathrm{Im}A^l$,这时有 $\boldsymbol{z}$ 使得 $\boldsymbol{x}=A^l\boldsymbol{z}$,于是,$YA\boldsymbol{x}=YA^{l+1}\boldsymbol{z}=A^l\boldsymbol{z}=\boldsymbol{x}$.

按定义 5,这表明 Y 为 A 的一个广义左逆. □

我们指出,若 $A\in \boldsymbol{M}_n(\mathbb{C})$ 有 $\text{index}(A)=k$,则定理 6 蕴涵 A 的每一个广义左逆恰为 A 的$\{6^k\}$-逆(见 5.2.1 小节方程(6^k)),因而 A^{D} 为 A 的一个广义左逆. 应用对于广义右逆的类似结果也能推出,A^{D} 为 A 的一个广义右逆.

习题 5.2.2

1. 设 $A\in \boldsymbol{M}_n(\mathbb{C})$且 $0\neq\lambda\in\mathbb{C}$. 试证:若对某个正整数 p,$(A-\lambda I)^p\boldsymbol{x}=\boldsymbol{0}$,则 $\boldsymbol{x}\in \text{Im}A^l$, $\forall\, l\geqslant 0$.

2. 设 $A\in \boldsymbol{M}_n(\mathbb{C})$有 $\text{index}(A)\leqslant 1$. 试证:对任意 $\lambda\in\mathbb{C}$ 与 $\boldsymbol{x}\in\mathbb{C}^n$,$(A-\lambda I)^p\boldsymbol{x}=\boldsymbol{0}$,$(A-\lambda I)^{p-1}\boldsymbol{x}\neq\boldsymbol{0}$ 当且仅当$(A^{\#}-\lambda^{\dagger}I)^p\boldsymbol{x}=\boldsymbol{0}$,$(A^{\#}-\lambda^{\dagger}I)^{p-1}\boldsymbol{x}\neq\boldsymbol{0}$. 并且 $A^{\#}$ 是 $A\{1\}\cup A\{2\}$ 中具有这种性质的唯一元素.

3. 证明命题 3,并证明类似于定理 6 对于广义右逆的结果.

4. 设 $A\in \boldsymbol{M}_n(\mathbb{C})$. 试证:$\text{index}(A)\leqslant 1$ 当且仅当

$$\lim_{\lambda\to 0}(\lambda I_n+A)^{-1}A$$

存在. 此时,

$$\lim_{\lambda\to 0}(\lambda I_n+A)^{-1}A = AA^{\#}.$$

5. 设 $A\in \boldsymbol{M}_n(\mathbb{C})$. 试证:$A$ 是值域 Hermite 矩阵当且仅当$\lim\limits_{\lambda\to 0}(\lambda I_n+A)^{-1}P_{\text{Im}A}=A^{\dagger}$.

6. 设 $O\neq A\in \boldsymbol{M}_{m,n}(\mathbb{C})$. 试证:

$$\lim_{\lambda\to 0}(\lambda I_n+A^*A)^{-1}A^* = A^{\dagger}.$$

7. 试证:$A\in \boldsymbol{M}_n(\mathbb{C})$有唯一分解 $A=B+N$ 使得 $\text{index}(B)\leqslant 1$,$N$ 为幂零矩阵,且 $BN=NB=O$. 此时 $B=(A^{\mathrm{D}})^{\#}$.

8. 设$\{A_j\}_{j=1}^{\infty}\subset \boldsymbol{M}_{m,n}(\mathbb{C})$,$A\in \boldsymbol{M}_{m,n}(\mathbb{C})$,且$\lim\limits_{j\to\infty}A_j=A$. 试证:$\lim\limits_{j\to\infty}A_j^{\dagger}=A^{\dagger}$ 当且仅当存在 j_0 使得

$$\text{rank}A_j = \text{rank}A, \qquad \forall\, j\geqslant j_0.$$

9. 应用题 8 证明:

(1) 若 $\mathscr{S}$与 $\mathscr{T}$为$\mathbb{C}^n$ 的子空间,$\dim\mathscr{S}>\dim\mathscr{T}>0$,则对 $\mathscr{T}$的任意补空间 $\mathscr{K}$,$\mathscr{S}\cap\mathscr{K}\neq\{\boldsymbol{0}\}$.

(2) 设 $P_j(j=1,2,\cdots)$与 P 为幂等矩阵,且$\lim\limits_{j\to\infty}P_j=P$. 则存在 j_0 使得 $\text{rank}P_j=\text{rank}P$, $\forall\, j\geqslant j_0$.

(3) 设 $A_j(j=1,2,\cdots)$,$A\in \boldsymbol{M}_n(\mathbb{C})$,且$\lim\limits_{j\to\infty}A_j=A$. 若 $k_j=\text{index}(A_j)=\text{index}(A)=k$, $\forall\, j\geqslant j_0$,且 $\text{rank}A_j^k=\text{rank}A^k$, $\forall\, j\geqslant j_0$,则$\lim\limits_{j\to\infty}A_j^{\mathrm{D}}=A^{\mathrm{D}}$.

5.2.3 矩阵的广义逆正性与单调性

在本小节我们只研究实矩阵与实向量. 元素均为非负实数(正数)的矩阵 $A\in \boldsymbol{M}_{m,n}(\mathbb{C})$称为**非负矩阵(正矩阵)**,记为 $A\geqslant O(A>O)$,(注意,在本书中有时也用 $A\geqslant O$或 $A>O$ 分别表示 Hermite 正半定或 Hermite 正定矩阵. 只要顾及上下文内容,一般是容易区分它们的)分量均为非负实数(正数)的向量 $\boldsymbol{x}\in\mathbb{R}^n$ 称为**非负(正)n 维向量**,记为 $\boldsymbol{x}\geqslant\boldsymbol{0}(\boldsymbol{x}>\boldsymbol{0})$. 非负 n 维向量的集合记为$\mathbb{R}_+^n$. 若 $\boldsymbol{x}-\boldsymbol{y}\geqslant\boldsymbol{0}$ 且 $\boldsymbol{x}$,

$\boldsymbol{y}\in\mathbb{R}^n$,则记 $\boldsymbol{x}\geqslant\boldsymbol{y}$.

定义 1　设 $A\in\boldsymbol{M}_n(\mathbb{R})$. 若 A 非奇异且 $A^{-1}\geqslant O$,则 A 叫作**逆正的**. 若 $A\boldsymbol{x}\geqslant\boldsymbol{0}$ 与 $\boldsymbol{x}\in\mathbb{R}^n$ 蕴涵 $\boldsymbol{x}\geqslant\boldsymbol{0}$,则 A 叫作**单调的**.

显然,$A\in\boldsymbol{M}_n(\mathbb{R})$为单调的当且仅当

$$A\boldsymbol{x}\geqslant A\boldsymbol{y}\Rightarrow\boldsymbol{x}\geqslant\boldsymbol{y}.$$

$A\in\boldsymbol{M}_n(\mathbb{R})$为逆正的当且仅当线性方程组 $A\boldsymbol{x}=\boldsymbol{b}$ 对任意 $\boldsymbol{b}\geqslant\boldsymbol{0}$ 有解 $\boldsymbol{x}=A^{-1}\boldsymbol{b}\geqslant\boldsymbol{0}$,即有非负解 $\boldsymbol{x}$. 下面定理表明,实方阵的逆正性等价于单调性.

定理 2　设 $A\in\boldsymbol{M}_n(\mathbb{R})$. 则 A 为逆正的当且仅当 A 为单调的.

证明　对非奇异矩阵 A 来说,显然有

$$\mathbb{R}_+^n\subset A\mathbb{R}_+^n\Leftrightarrow A^{-1}\mathbb{R}_+^n\subset\mathbb{R}_+^n.$$

但上式左端命题等价于 A 的单调性,(因为当左端命题成立时,$A\boldsymbol{x}\geqslant\boldsymbol{0}$ 与 $\boldsymbol{x}\in\mathbb{R}^n$ 蕴涵存在 $\boldsymbol{y}\in\mathbb{R}_+^n$ 使得 $A\boldsymbol{x}=A\boldsymbol{y}$,因而由 A 的非奇异性推出 $\boldsymbol{x}=\boldsymbol{y}\geqslant\boldsymbol{0}$. 反之,从 A 的单调性看出,若 $\boldsymbol{x}\geqslant\boldsymbol{0}$,则 $A(A^{-1}\boldsymbol{x})\geqslant\boldsymbol{0}$,因而 $\boldsymbol{y}=A^{-1}\boldsymbol{x}\geqslant\boldsymbol{0}$,这表明 $\boldsymbol{x}=A\boldsymbol{y}\in A\mathbb{R}_+^n$)右端命题等价于 A 的逆正性. 余下只要证:A 单调蕴涵 A^{-1} 存在. 事实上,当 A 单调时,$A\boldsymbol{x}=\boldsymbol{0}$ 与 $A(-\boldsymbol{x})=\boldsymbol{0}$ 蕴涵 $\boldsymbol{x}\geqslant\boldsymbol{0}$ 与 $-\boldsymbol{x}\geqslant\boldsymbol{0}$,因而 $A\boldsymbol{x}=\boldsymbol{0}$ 蕴涵 $\boldsymbol{x}=\boldsymbol{0}$,即 A 为非奇异的. □

我们将在后面第八章 8.1.2 小节中继续讨论单调(逆正)方阵的其他特征. 在这里主要讨论矩阵 $A\in\boldsymbol{M}_{m,n}(\mathbb{R})$的广义单调性与广义逆正性的概念与特征.

定义 3　设 $A\in\boldsymbol{M}_{m,n}(\mathbb{R})$. 当 λ 为集合$\{1,2,3,4\}$的一个不空子集时,A 叫作 **λ-单调的**,假如 A 有一个非负的 λ-逆. 特殊地,A 叫作**半单调的**,假如 $A^{\dagger}\geqslant O$. 方阵 $A\in\boldsymbol{M}_n(\mathbb{R})$叫作**群单调的**,假如 $A^{\#}$ 存在且 $A^{\#}\geqslant O$;方阵 $A\in\boldsymbol{M}_n(\mathbb{R})$叫作 ***D*-单调的**,假如 $A^{\mathrm{D}}\geqslant O$.

定义 4　设 $A\in\boldsymbol{M}_{m,n}(\mathbb{R})$且 $\mathscr{S}$为$\mathbb{R}^n$ 的子集. A 叫作**在$\mathscr{S}$上非负的**,假如 $\boldsymbol{x}\in\mathbb{R}_+^n\cap\mathscr{S}$蕴涵 $A\boldsymbol{x}\geqslant\boldsymbol{0}$;$A$ 叫作**在$\mathscr{S}$上单调的**,假如 $A\boldsymbol{x}\geqslant\boldsymbol{0}$ 与 $\boldsymbol{x}\in\mathscr{S}$蕴涵 $\boldsymbol{x}\geqslant\boldsymbol{0}$.

先考虑 D-单调性的特征.

定理 5　设 $A\in\boldsymbol{M}_n(\mathbb{R})$. 则 A 为 D-单调的当且仅当对某个整数 $k\geqslant0$,

$$A\boldsymbol{x}\in\mathbb{R}_+^n+\mathrm{Ker}A^k\text{ 与 }\boldsymbol{x}\in\mathrm{Im}A^k\Rightarrow\boldsymbol{x}\geqslant\boldsymbol{0}.\tag{1}$$

并且,若(1)成立,则 $k\geqslant\mathrm{index}(A)$.

证明　先证(1)式蕴涵 $k\geqslant\mathrm{index}(A)$与 $A^{\mathrm{D}}\geqslant O$. 设 $A^{k+1}\boldsymbol{x}=\boldsymbol{0}$,这时,$A^k\boldsymbol{x}\in\mathrm{Im}A^k$ 与 $A(A^k\boldsymbol{x})=\boldsymbol{0}$,故(1)式蕴涵 $A^k\boldsymbol{x}=\boldsymbol{0}$. 这表明 $\mathrm{Ker}A^{k+1}\subset\mathrm{Ker}A^k$,因而 $k\geqslant\mathrm{index}(A)$. 于是,$\mathbb{R}^n=\mathrm{Im}A^k\dot{+}\mathrm{Ker}A^k$(见 5.2.1 小节(10)式). 现设 $\boldsymbol{w}\geqslant\boldsymbol{0}$ 有分解 $\boldsymbol{w}=\boldsymbol{z}-\boldsymbol{y}$ 使得 $\boldsymbol{z}\in\mathrm{Im}A^k$ 与 $\boldsymbol{y}\in\mathrm{Ker}A^k$. 由于 $\mathrm{Im}A^k=\mathrm{Im}A^{k+1}$,故有 $\boldsymbol{x}\in\mathrm{Im}A^k$ 使得 $A\boldsymbol{x}=\boldsymbol{z}=\boldsymbol{w}+\boldsymbol{y}$,这时,(1)式保证 $\boldsymbol{x}\geqslant\boldsymbol{0}$. 因此,按习题 5.2.1 题 2 有 $A^{\mathrm{D}}\boldsymbol{w}=A^{\mathrm{D}}A\boldsymbol{x}-A^{\mathrm{D}}\boldsymbol{y}=\boldsymbol{x}\geqslant\boldsymbol{0}$, $\forall\,\boldsymbol{w}\geqslant\boldsymbol{0}$,即有 $A^{\mathrm{D}}\geqslant O$.

反之,设 $A^{\mathrm{D}}\geqslant O$ 且 $k\geqslant\mathrm{index}(A)$. 令 $A\boldsymbol{x}=\boldsymbol{w}+\boldsymbol{y}$,使得 $\boldsymbol{w}\geqslant\boldsymbol{0}$,$\boldsymbol{y}\in\mathrm{Ker}A^k$ 与 $\boldsymbol{x}\in$

$\mathrm{Im}A^k$. 由此推出 $A^D\boldsymbol{w}\geqslant\boldsymbol{0}, A^D\boldsymbol{y}=\boldsymbol{0}$ 与 $A^DA\boldsymbol{x}=\boldsymbol{x}$，因而 $\boldsymbol{x}=A^DA\boldsymbol{x}=A^D\boldsymbol{w}+A^D\boldsymbol{y}=A^D\boldsymbol{w}\geqslant\boldsymbol{0}$. □

特别地，若 $1\geqslant\mathrm{index}(A)$，则有 $A^D=A^{\#}$，因而从定理 5 直接得到 A 为群单调的充分与必要条件.

定理 6 设 $A\in\boldsymbol{M}_n(R)$. 则 A 为群单调的当且仅当

$$A\boldsymbol{x}\in\mathbb{R}_+^n+\mathrm{Ker}A \text{ 与 } \boldsymbol{x}\in\mathrm{Im}A\Rightarrow\boldsymbol{x}\geqslant\boldsymbol{0}. \tag{2}$$

其次，对于 $A\in\boldsymbol{M}_n(\mathbb{R})$ 的 Moore-Penrose 逆，类似地可以得到 A 半单调的充分与必要条件(其证明留给读者，见习题 6).

定理 7 设 $A\in\boldsymbol{M}_n(\mathbb{R})$. 则 A 半单调当且仅当

$$A\boldsymbol{x}\in\mathbb{R}_+^n+\mathrm{Ker}A^{\mathrm{T}} \text{ 与 } \boldsymbol{x}\in\mathrm{Im}A^{\mathrm{T}}\Rightarrow\boldsymbol{x}\geqslant\boldsymbol{0}. \tag{3}$$

我们指出，当 $A\in\boldsymbol{M}_n(\mathbb{R})$ 为值域 Hermite 矩阵，即有 $\mathrm{Im}A^{\mathrm{T}}=\mathrm{Im}A$ 时，条件(2)与(3)重合.

最后考虑 $A\in\boldsymbol{M}_n(R)$ 的**广义左逆正性**，即 A 有一个广义左逆 Y 满足 $Y\geqslant O$. A 的**广义右逆正性**的类似结果的证明留给读者. 为此，先证一个引理，它与 $\mathbb{R}_+^n$ 上线性不等式的理论密切相关.

引理 8 设 $C\in\boldsymbol{M}_n(\mathbb{R})$ 与 $\boldsymbol{b}\in\mathbb{R}^n$. 则下列两命题彼此等价：

(1) $C\boldsymbol{y}=\boldsymbol{b}$ 与 $\boldsymbol{y}\in\mathbb{R}_+^n$ 是相容的，亦即 $C\boldsymbol{y}=\boldsymbol{b}$ 有非负向量解 $\boldsymbol{y}$.

(2) $C^{\mathrm{T}}\boldsymbol{x}\geqslant\boldsymbol{0}$ 与 $\boldsymbol{x}\in\mathbb{R}^n\Rightarrow\boldsymbol{x}^{\mathrm{T}}\boldsymbol{b}\geqslant 0$.

证明 (1)⇒(2)是显然的，因为若 $C\boldsymbol{y}=\boldsymbol{b}$ 有解 $\boldsymbol{y}\geqslant\boldsymbol{0}$，且 $C^{\mathrm{T}}\boldsymbol{x}\geqslant\boldsymbol{0}$ 与 $\boldsymbol{x}\in\mathbb{R}^n$，则 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{b}=\boldsymbol{x}^{\mathrm{T}}C\boldsymbol{y}=(C^{\mathrm{T}}\boldsymbol{x})^{\mathrm{T}}\boldsymbol{y}\geqslant 0$.

(2)⇒(1)：设 $C=[\boldsymbol{c}_1\quad\boldsymbol{c}_2\quad\cdots\quad\boldsymbol{c}_n]$，并令 $\mathscr{S}$ 及其**对偶集** $\mathscr{S}^*$ 如下：

$$\mathscr{S}\equiv\{\alpha_1\boldsymbol{c}_1+\alpha_2\boldsymbol{c}_2+\cdots+\alpha_n\boldsymbol{c}_n:(\alpha_1,\alpha_2,\cdots,\alpha_n)^{\mathrm{T}}\geqslant\boldsymbol{0}\},$$

$$\mathscr{S}^*\equiv\{\boldsymbol{x}\in\mathbb{R}^n:\boldsymbol{y}\in\mathscr{S}\Rightarrow\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}\geqslant 0\}.$$

当(2)成立时，$C^{\mathrm{T}}\boldsymbol{x}\geqslant\boldsymbol{0}$ 与 $\boldsymbol{x}\in\mathbb{R}^n$ 相当于：$\boldsymbol{x}\in\mathbb{R}^n$ 与 $\boldsymbol{x}^{\mathrm{T}}C\boldsymbol{\alpha}\geqslant 0$, $\forall\boldsymbol{\alpha}=(\alpha_1,\cdots,\alpha_n)^{\mathrm{T}}\geqslant\boldsymbol{0}$. 按 $\mathscr{S}$ 与 $\mathscr{S}^*$ 的定义，此等价于 $\boldsymbol{x}\in\mathscr{S}^*$. 而 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{b}\geqslant 0$, $\forall\boldsymbol{x}\in\mathscr{S}^*$ 等价于 $\boldsymbol{b}\in\mathscr{S}^{**}\equiv\{\boldsymbol{w}\in\mathbb{R}^n:\boldsymbol{x}\in\mathscr{S}^*\Rightarrow\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}\geqslant 0\}$ (φ^* 的对偶集). 但容易证明，$\mathscr{S}$ 是 $\mathbb{R}^n$ 内闭凸锥，且 $\mathscr{S}^{**}\subset\overline{\mathscr{S}}=\mathscr{S}$. 事实上，设 $\boldsymbol{x}\notin\mathscr{S}$，则由第二章 2.3.1 小节定理 4 看出，存在 $\boldsymbol{0}\neq\boldsymbol{u}\in\mathbb{R}^n$ 使得 $\boldsymbol{u}^{\mathrm{T}}\boldsymbol{y}\geqslant 0$ ($\forall\boldsymbol{y}\in\mathscr{S}$) 与 $\boldsymbol{u}^{\mathrm{T}}\boldsymbol{x}<0$. 前者表明 $\boldsymbol{u}\in\mathscr{S}^*$，因而后者表明 $\boldsymbol{x}\notin\mathscr{S}^{**}$. 因此，命题(2)蕴涵 $\boldsymbol{b}\in\mathscr{S}$，这等价于 $C\boldsymbol{y}=\boldsymbol{b}$ 有非负解 $\boldsymbol{y}\geqslant\boldsymbol{0}$. □

定理 9 设 $A\in\boldsymbol{M}_n(\mathbb{R})$ 与 $\mathscr{T}=\bigcap\limits_{m=0}^{\infty}\mathrm{Im}A^m$. 则下列诸命题彼此等价：

(1) A 为广义左逆正的，即 A 有一个广义左逆 $Y\geqslant O$.

(2) A 有一个广义左逆 Y，它在 $\mathscr{T}$ 上非负，即有 $Y(\mathbb{R}_+^n\cap\mathscr{T})\subset\mathbb{R}_+^n$.

(3) A 的每一个广义左逆在 $\mathscr{T}$ 上非负(因而特别地 A^D 在 $\mathscr{T}$ 上非负).

(4) A 在 $\mathscr{T}$ 上单调.

证明 (1)⇒(2)显然.(2)⇒(3):设 $L\in \boldsymbol{M}_n(R)$ 为 A 的任一广义左逆,且 $\boldsymbol{x}\in \mathbb{R}_+^n\cap\mathscr{T}$. 这时存在 $\boldsymbol{z}\in\mathbb{R}^n$ 使得 $\boldsymbol{x}=A^{k+1}\boldsymbol{z}$,其中 $k=\text{index}(A)$,于是,按 5.2.2 小节定理 6(2),

$$A^k\boldsymbol{z}=YA^{k+1}\boldsymbol{z}=Y\boldsymbol{x}\geqslant \boldsymbol{0},$$

因为 Y 在 $\mathscr{T}$ 上非负.类似地,$L\boldsymbol{x}=LA^{k+1}\boldsymbol{z}=A^k\boldsymbol{z}\geqslant\boldsymbol{0}$,即 L 在 $\mathscr{T}$ 上非负.

(3)⇒(4):设(3)成立且 $A\boldsymbol{x}\geqslant\boldsymbol{0}$ 与 $\boldsymbol{x}\in\mathscr{T}$. 令 $L\in\boldsymbol{M}_n(\mathbb{R})$ 为 A 的某个广义左逆,按定义,$\boldsymbol{x}=LA\boldsymbol{x}$. 但由于 $A\boldsymbol{x}\in\mathscr{T}$,故根据(3)有 $\boldsymbol{x}=L(A\boldsymbol{x})\geqslant\boldsymbol{0}$.

(4)⇒(1):设 $\text{index}(A)=k$,且 $(A^k)_{(i)}$ 表示 A^k 的第 i 个行向量,$i=1,2,\cdots,n$. 若 $A^{k+1}\boldsymbol{x}\geqslant\boldsymbol{0}$,$\boldsymbol{x}\in\mathbb{R}^n$,则由于 $\mathbb{R}^n=\mathscr{T}\dotplus\text{Ker}A^k$,故有 $\boldsymbol{x}=\boldsymbol{u}+\boldsymbol{v}$,其中 $\boldsymbol{u}\in\mathscr{T}$ 与 $\boldsymbol{v}\in\text{Ker}A^k$,因而

$$A(A^k\boldsymbol{u})=A^{k+1}\boldsymbol{u}=A^{k+1}\boldsymbol{x}\geqslant\boldsymbol{0}\quad 与\ A^k\boldsymbol{u}\in\mathscr{T}.$$

根据(4),$A^k\boldsymbol{u}\geqslant\boldsymbol{0}$. 但是,$(A^k)_{(i)}\boldsymbol{x}=(A^k)_{(i)}\boldsymbol{u}+(A^k)_{(i)}\boldsymbol{v}=(A^k)_{(i)}\boldsymbol{u}\geqslant 0$,因此,$A^{k+1}\boldsymbol{x}\geqslant\boldsymbol{0}$ 与 $\boldsymbol{x}\in\mathbb{R}^n$ 蕴涵 $(A^k)_{(i)}\boldsymbol{x}\geqslant 0$,$i=1,2,\cdots,n$. 应用引理 8(取 $A^{k+1}=C^{\mathrm{T}}$ 与 $(A^k)_{(i)}=\boldsymbol{b}^{\mathrm{T}}$)我们得到,$\boldsymbol{y}^{\mathrm{T}}A^{k+1}=(A^k)_{(i)}$ 有解 $\boldsymbol{y}\geqslant\boldsymbol{0}$,$i=1,2,\cdots,n$,此表明 $YA^{k+1}=A^k$ 与 $Y\geqslant O$ 是相容的. 按 5.2.2 小节定理 6,此非负 n 阶矩阵 Y 为 A 的一个广义左逆. □

假若 $A\in\boldsymbol{M}_n(\mathbb{R})$ 有 $\text{index}(A)\leqslant 1$,则在前一定理中,$\mathscr{T}=\text{Im}A$,且这时 $A^{\#}=A^{\mathrm{D}}$. 因此,由定理 9 立即得出

推论 10 设 $A\in\boldsymbol{M}_n(\mathbb{R})$ 有 $\text{index}(A)\leqslant 1$. 则 $A^{\#}$ 在 $\text{Im}A$ 上非负当且仅当 A 在 $\text{Im}A$ 上单调.

同理可得定理 2 的如下推广.

推论 11 设 $A\in\boldsymbol{M}_n(\mathbb{R})$ 有 $\text{index}(A)=k$. 则 A^{D} 在 $\text{Im}A^k$ 上非负当且仅当 A 在 $\text{Im}A^k$ 上单调.

习题 5.2.3

1. 设 $A\in\boldsymbol{M}_n(\mathbb{R})$. 若对 $\mathbb{R}^n$ 中 $\text{Ker}A$ 的每一个补空间 $\mathscr{T}$,A 在 $\mathscr{T}$ 上单调,则称 A 为**几乎单调的**. 试证:当 $\text{Ker}A\neq\{\boldsymbol{0}\}$ 时,下列命题彼此等价:

(1) A 为几乎单调的.

(2) $A\boldsymbol{x}\geqslant\boldsymbol{0}$ 蕴涵 $A\boldsymbol{x}=\boldsymbol{0}$.

2. 设 $A\in\boldsymbol{M}_n(\mathbb{R})$. 试证:$A$ 在 $\text{Im}A^{\mathrm{T}}$ 上单调当且仅当它是{1,4}-单调的.

3. 设 $A\in\boldsymbol{M}_{m,n}(\mathbb{R})$,$\mathscr{S}$ 为 $\text{Im}A$ 的补空间,$\mathscr{T}$ 为 $\text{Ker}A$ 的补空间. 令 $A_{\mathscr{T},\mathscr{S}}^{[1,2]}$ 表示以 $\mathscr{T}$ 为值域与以 $\mathscr{S}$ 为零空间的 A 的唯一的{1,2}-逆(见 5.1.2 小节推论 6). 试证:若 P 与 Q 分别为 $\mathbb{R}^n$ 与 $\mathbb{R}^m$ 的不空子集,则下列诸命题彼此等价:

(1) $A_{\mathscr{T},\mathscr{S}}^{[1,2]}Q\subset P$.

(2) $A\boldsymbol{x}\in AA_{\mathscr{T},\mathscr{S}}^{[1,2]}Q\Rightarrow A_{\mathscr{T},\mathscr{S}}^{[1,2]}A\boldsymbol{x}\in P$.

(3) $A\boldsymbol{x}\in Q\dotplus\mathscr{S}\Rightarrow A_{\mathscr{T},\mathscr{S}}^{[1,2]}A\boldsymbol{x}\in P$.

(4) $Ax\in AA_{\mathcal{T},\mathcal{S}}^{[1,2]}Q$ 与 $x\in\mathcal{T}\Rightarrow x\in P$.

(5) $Ax\in Q+\mathcal{S}$与 $x\in\mathcal{T}\Rightarrow x\in P$.

4. 沿用题 3 中符号. 试证:$A_{\mathcal{T},\mathcal{S}}^{[1,2]}(Q\cap \mathrm{Im}A)\subset P$ 当且仅当 $Ax\in Q$ 与 $x\in\mathcal{T}$ 蕴涵 $x\in P$. 并且,$A_{\mathcal{T},\mathcal{S}}^{[1,2]}Q\subset P+\mathrm{Ker}A$ 当且仅当 $Q\subset AP+\mathcal{S}$.

5. 沿用题 3 中的符号. 试证下列诸命题等价:

(1) $A_{\mathcal{T},\mathcal{S}}^{[1,2]}(Q\cap \mathrm{Im}A)\subset P+\mathrm{Ker}A$.

(2) $Ax\in Q\Rightarrow x\in P+\mathrm{Ker}A$.

(3) $Ax\in Q$ 与 $x\in\mathcal{T}\Rightarrow x\in P+\mathrm{Ker}A$.

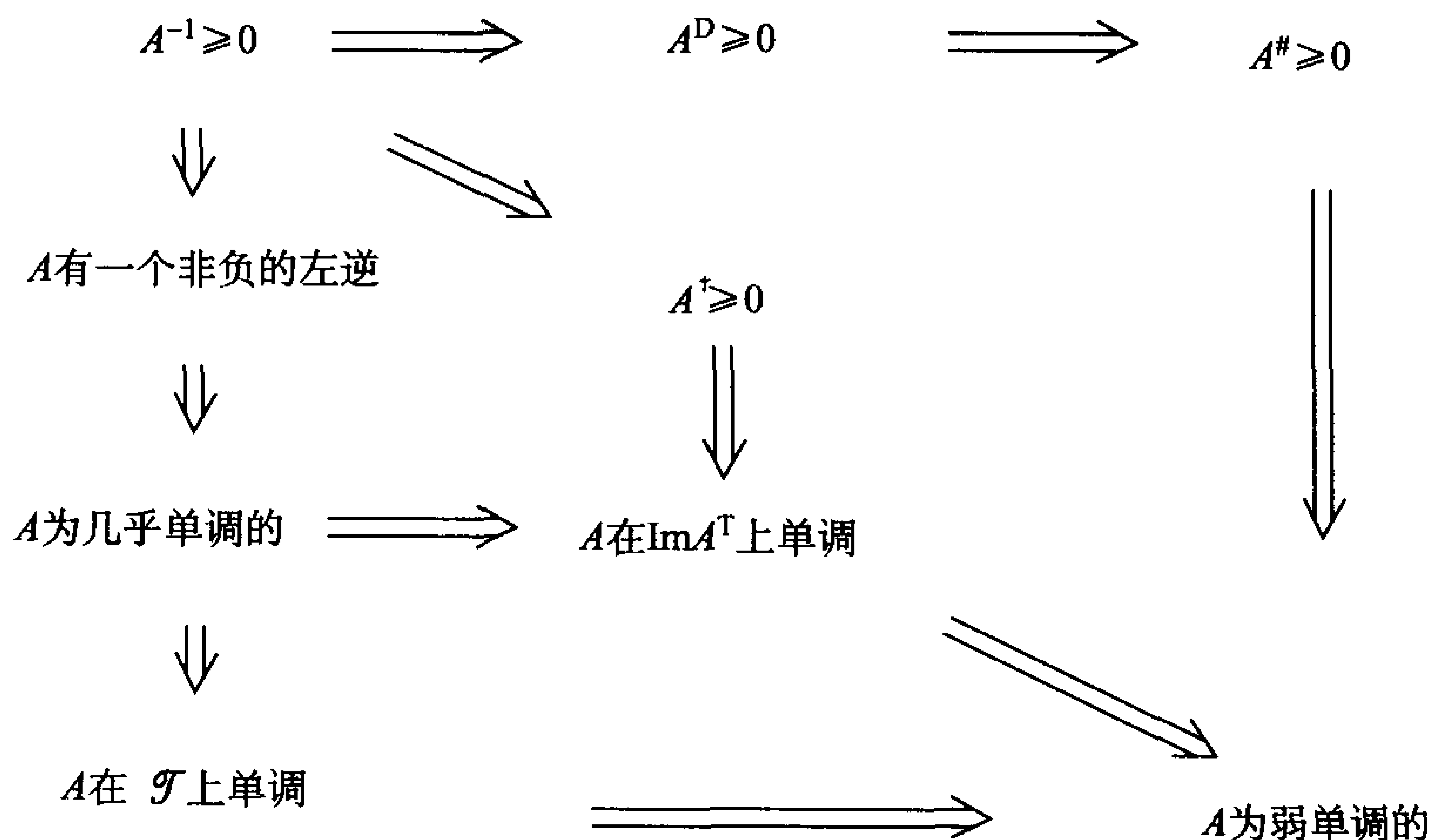
(4) $Q\cap \mathrm{Im}A\subset AP$.

(5) $Q\cap \mathrm{Im}A\subset AP+\mathcal{S}$.

6. 设 $A\in \boldsymbol{M}_n(\mathbb{R})$. 试证:$A^{\dagger}\geqslant O$ 当且仅当

$$Ax\in\mathbb{R}_+^n+\mathrm{Ker}A^{\mathrm{T}}\quad 与\quad x\in \mathrm{Im}A^{\mathrm{T}}\Rightarrow \boldsymbol{x}\geqslant\boldsymbol{0}.$$

7. 设 $A\in \boldsymbol{M}_n(\mathbb{R})$且 $\mathcal{T}$为 KerA 的补空间. 验证下图中各逻辑包含关系:



其中,A 为弱单调的意指 $Ax\geqslant\boldsymbol{0}$ 必蕴涵 $x\in\mathbb{R}_+^n+\mathrm{Ker}A$;$A$ 为几乎单调的涵义见题 1.

8. 设 $C\in \boldsymbol{M}_{m,n}(\mathbb{R})$与 $\boldsymbol{b}\in\mathbb{R}^m$. 试证下列两命题彼此等价:

(1) $C\boldsymbol{y}=\boldsymbol{b}$ 有非负向量解 $\boldsymbol{y}$.

(2) $C^{\mathrm{T}}\boldsymbol{x}\geqslant\boldsymbol{0}$ 与 $\boldsymbol{x}\in\mathbb{R}^m\Rightarrow \boldsymbol{x}^{\mathrm{T}}\boldsymbol{b}\geqslant 0$.

参考文献

[1] Ben-Israel A, Greville T N E. Generalized Inverses: Theory and Applications. 2nd ed. New York: Springer-Verlag, 2003

[2] Nashed M. Generalized Inverses and Applications. New York: Academic Press, 1976

[3] Boullion T L, Odell P L. Generalized Inverse Matrices. New York: Wiley-Interscience, 1971

[4] Rao C R, Mitra S. Generalized Inverse of Matrices with Applications in Statistics. New York: Wiley, 1971

[5] Campbell S L. Recent applications of generalized inverses. Vol. 66, Research Notes in Math. Boston: Pitman, 1982

[6] Penrose R. A generalized inverse for matrices. Proc. Cambridge Philos. Soc., 1955(51), 406～413

[7] Rhode C A. Some results on generalized inverses. SIAM Rev., 1966, 8(2): 201～205

[8] Cline R E, Greville T N E. A Drazin inverse for rectangular matrices. Linear Algebra Appl., 1980(29), 53～62

[9] Berman A, Plemmons R J. Nonnegative Matrices in the Mathematical Sciences. New York: Academic Press, 1979, Reprinted and updated, SIAM Press, Philadelphia, 1994

[10] Berman A, Plemmons R J. Matrix group monotonicity. Proc. Amer. Math. Soc., 1974(46), 355～359

[11] Berman A, Plemmons R J. Eight types of matrix monotonicity. Linear Algebra Appl., 1976 (13), 115～123

[12] Pye Wallace C. Nonnegative Drazin inverses. Linear Algebra Appl., 1980(30), 149～153

[13] Carlson D. Generalizations of matrix mononicity. Linear Algebra Appl., 1976 (13), 125～131

[14] Campbell S L, Meyer C D J. Generalized Inverses of Linear Transformations. London: Pitman, 1979

第六章　特征值的定位与扰动

本章前两节研究复方阵之特征值的位置估计问题，第三节讨论矩阵特征值的连续性与扰动问题. 这些问题都具有理论上与应用上的重要价值.

6.1　矩阵非奇异性定理与排除定理

给定矩阵 $A\in \boldsymbol{M}_n(\mathbb{C})$，我们将给出 A 的特征值在复平面位置信息的任意定理称为**定位定理**. 复平面内至少含有 A 的一个特征值的区域称为 A 的**包含区域**，给出包含区域的定理称为**包含定理**；复平面内不含有 A 的任何特征值的区域称为 A 的**排除区域**，给出排除区域的定理称为**排除定理**. 这两种区域可以是连通的或者非连通的. 可以是有界的或者无界的. 这两种定理都是定位定理.

我们自然希望得到尽可能小的包含区域与尽可能大的排除区域，并且希望用尽可能少的关于矩阵的信息与计算得到这两种区域. 这些结果给出的矩阵特征值的估计可能是比较粗糙的，但有时候，譬如判定矩阵稳定性或矩阵级数的收敛性，即便是这样粗糙的信息也就足够了.

存在一些简单的排除定理，例如，Hermite 矩阵的特征值只能在实轴上（即复平面内不含实轴的区域为其排除区域）；斜 Hermite 矩阵的特征值只能在虚轴上；酉矩阵的特征值只能在单位圆周上，等等. 另外，在第二章 2.2.2 小节定理 4 中指出，对 $\boldsymbol{M}_n(\mathbb{C})$ 上任意矩阵范数 $\|\cdot\|$，$|\lambda|>\|A\|$ 为 A 的一个排除区域.

排除定理总可以视为断定某些矩阵为非奇异的定理. 这是因为若 α 属于 $A\in \boldsymbol{M}_n(\mathbb{C})$ 的某个排除区域，则 $\alpha\notin\sigma(A)$，因而矩阵 $\alpha I-A$ 非奇异. 反之，许多断定某些矩阵为非奇异的定理经过适当解释后均可确定出相应的排除定理. 这些非奇异性定理的叙述与推导都比较简单，它们主要借助于矩阵元素的某些简单函数或有关矩阵的元素的某些简单函数来判定矩阵的非奇异性.

6.1.1　严格对角占优矩阵与 Gerschgorin 圆盘定理

矩阵非奇异性定理种类繁多，形式各异，但其中最著名也是最古老的首推 **Lévy-Desplangues 定理**（也称为 **Hadamard 定理**）. 为此，先引入对角占优矩阵的概念.

定义 1　设 $A=(a_{ij})\in \boldsymbol{M}_n(\mathbb{C})$ 且 $n\geqslant 2$. 如果

$$|a_{ii}| \geqslant R_i(A) \equiv \sum_{j\neq i} |a_{ij}|, \qquad i=1,2,\cdots,n, \tag{1}$$

这里，$\sum_{j\neq i}|a_{ij}|$ 表示 $\sum_{j=1,j\neq i}^{n}|a_{ij}|$，则称 A 为(行) **对角占优的**. 如果

$$|a_{ii}| > R_i(A), \qquad i=1,2,\cdots,n. \tag{2}$$

则称 A 为(行)**严格对角占优的**. 类似地，可以定义(列)**对角占优**矩阵与(列)**严格对角占优**矩阵 A，只要(1)与(2)两式中 $R_i(A)$ 用

$$C_i(A) \equiv \sum_{j\neq i} |a_{ji}|, \qquad i=1,2,\cdots,n \tag{3}$$

代替即可.

今后如无特殊声明，对角占优或严格对角占优总是对行而言的. 这两种关于行与列的定义通过矩阵转置可以互相变换，例如，A 行对角占优当且仅当 A^{T} 列对角占优.

对角占优矩阵可能是奇异的，如 $A=\begin{bmatrix}1 & -1\\ -1 & 1\end{bmatrix}$. 但是，下面定理指出，严格对角占优矩阵必定是非奇异的.

定理 2(Lévy-Desplanques)　若 $A=(a_{ij})\in \boldsymbol{M}_n(\mathbb{C})$ $(n\geqslant 2)$ 严格对角占优，则 $\det A\neq 0$.

证明　用反证法. 若 $\det A=0$，则齐次方程组 $A\boldsymbol{x}=\boldsymbol{0}$ 有非零解 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$. 令 $|x_r|=\max_{1\leqslant i\leqslant n}|x_i|$，则 $x_r\neq 0$，于是，$A\boldsymbol{x}=\boldsymbol{0}$ 的第 r 个方程蕴涵

$$|a_{rr}||x_r| \leqslant \sum_{j\neq r}|a_{rj}||x_j| \leqslant R_r(A)|x_r|, \tag{4}$$

由此推出，$|a_{rr}|\leqslant R_r(A)$，与(2)式矛盾. 因此，$\det A\neq 0$. □

由于 $A\in\boldsymbol{M}_n(\mathbb{C})$ 的非奇异性等价于它的相似矩阵 $S^{-1}AS$ 的非奇异性，故从定理 2 直接得出下面推论.

推论 3　若 $A=(a_{ij})\in\boldsymbol{M}_n(\mathbb{C})$ $(n\geqslant 2)$ 满足

$$|a_{ii}| > R_i^x(A) \equiv \frac{1}{x_i}\sum_{j\neq i}|a_{ij}|x_j, \qquad i=1,2,\cdots,n, \tag{5}$$

式中，$\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}>\boldsymbol{0}$，则 $\det A\neq 0$.

证明　令 $X=\mathrm{diag}[x_1,x_2,\cdots,x_n]$. 则(5)式断言 $X^{-1}AX$ 严格对角占优，因而 $\det(X^{-1}AX)=\det A\neq 0$. □

我们指出，条件(5)显然是条件(2)的推广. 通常将满足条件(5)的矩阵 A 称为**广义严格对角占优的**. 推论 3 断言，这样的矩阵总是非奇异的.

例如，$A=\begin{bmatrix}2 & -2\\ 0 & 1\end{bmatrix}$不是严格对角占优的，但存在 $X=\mathrm{diag}[2,1]$使得 $X^{-1}AX=\begin{bmatrix}2 & -1\\ 0 & 1\end{bmatrix}$严格对角占优. 于是，此 A 为广义严格对角占优的.

推论 4 设 $A=(a_{ij})\in\boldsymbol{M}_n(\mathbb{C})(n\geqslant2)$对角占优且它的所有对角元素 $a_{ii}\neq0$. 若(1)式中有 $n-1$ 个严格不等式成立,则有 $\det A\neq0$.

证明 设$|a_{rr}|=R_r(A)$;$|a_{ii}|>R_i(A)$,$\forall i\neq r$. 令 $x_r=1+\varepsilon(\varepsilon>0)$;$x_i=1$,$\forall i\neq r$. 则有

$$|a_{rr}|>\frac{1}{1+\varepsilon}R_r(A)=R_r^x(A),\quad \forall\varepsilon>0,$$

$$R_i(A)+\varepsilon|a_{ir}|=R_i^x(A),\qquad \forall i\neq r.$$

取 $\varepsilon>0$ 足够小使得$|a_{ii}|>R_i(A)+\varepsilon|a_{ir}|$,$\forall i\neq r$. 应用推论 3 便得 $\det A\neq0$. □

应用定理 2 与特征值连续性的结果(即矩阵特征值是矩阵元素的连续函数,见后面 6.3.1 小节定理 1),我们可以导出著名的 **Gerschgorin 圆盘定理**,它是矩阵的一种排除定理.

定理 5 设 $A=(a_{ij})\in\boldsymbol{M}_n(\mathbb{C})$且 $n\geqslant2$. 则 A 的任意特征值λ 都属于下列n 个圆盘 $\mathscr{D}_i$ 的并集$G(A)$:

$$\mathscr{D}_i=\{z\in\mathbb{C}:|z-a_{ii}|\leqslant R_i(A)\},\qquad i=1,2,\cdots,n. \tag{6}$$

若在(6)式所示圆盘中,有 k 个互相连通且与其余$n-k$ 个不相交,则 A 恰有 k 个特征值(计入重根)含在此 k 个圆盘组成的区域内.

证明 若有 $\lambda\in\sigma(A)\backslash G(A)$,则

$$|\lambda-a_{ii}|>R_i(A),\qquad i=1,2,\cdots,n,$$

即矩阵 $\lambda I-A$ 严格对角占优,因而按定理 2,$\det(\lambda I-A)\neq0$,与 $\lambda\in\sigma(A)$矛盾. 因此,$\lambda\in G(A)$,$\forall\lambda\in\sigma(A)$.

为证第二部分,定义矩阵族 $A_t=D+tB$,$\forall t\in[0,1]$,这里,$D=\mathrm{diag}[a_{11},a_{22},\cdots,a_{nn}]$,$B=A-D$. 则 $A_0=D$ 与 $A_1=A$. 令 $\mathscr{D}_i(t)$表示圆盘:$|z-a_{ii}|\leqslant tR_i(A)$,$i=1,2,\cdots,n$. 不失普遍性,假定(6)式中前 k 个圆盘组成连通区域 G,它与其余的 $n-k$ 个圆盘的并集 $\widetilde{G}$ 不相交. 记 $G(t)=\bigcup_{i=1}^{k}\mathscr{D}_i(t)$ 与 $\widetilde{G}(t)=\bigcup_{i=k+1}^{n}\mathscr{D}_i(t)$. 显然,$G(t)\subset G(1)=G$ 与 $\widetilde{G}(t)\subset\widetilde{G}(1)=\widetilde{G}$,$\forall t\in[0,1]$,因而对所有 $t\in[0,1]$,两个闭集 $G(t)$与 $\widetilde{G}(t)$不交. 特殊地,$G(0)$恰含 $A_0=D$ 的 k 个特征值:$a_{11},\cdots,a_{kk}$. 按第一部分结果,对所有 $t\in[0,1]$,A_t 的特征值全含在$G(t)\cup\widetilde{G}(t)$之中,但由于 $G(t)$与 $\widetilde{G}(t)$不交,当 t 由 0 增加到 1 时,按特征值连续性结果,A_t 的特征值不能从 $G(t)$跳到$\widetilde{G}(t)$,或从 $\widetilde{G}(t)$跳到 $G(t)$. 因此,$G(0)$恰有 $A_0=D$ 的 k 个特征值的事实保证了对所有 $t\in[0,1]$,$G(t)$必须恰含有 A_t 的 k 个特征值,因而 $G=G(1)$必须恰含有 A 的 k 个特征值. □

我们指出,上述定理的第二部分实际上断言,$G(A)$的每个分枝(即 $G(A)$的每个最大连通子集)内含有 A 的特征值的个数(计入重数)恰等于这个分枝内圆盘 $\mathscr{D}_i$ 的个数. 因此,每个孤立的圆盘 $\mathscr{D}_i$ 恰有 A 的一个单重的特征值.

将圆盘定理用于 A^{T},可得到类似的结果. 因此,

$$\sigma(A) \subset G(A) \cap G(A^{\mathrm{T}}). \tag{7}$$

应用推论 3，还可得到

$$\sigma(A) \subset \bigcap_D G(D^{-1}AD), \sigma(A) \subset \bigcap_D G(D^{-1}A^{\mathrm{T}}D), \tag{8}$$

式中，D 取遍所有 n 阶正对角矩阵.

这些结果都给出矩阵的排除区域，取它们的交一般可以得到较满意的特征值估计.

例 6　对于四阶复矩阵

$$A = \begin{bmatrix} 2 & -1/2 & 1/4 & -1/4 \\ 1/4 & 1+2\mathrm{i} & 0 & 1/4 \\ -1/2 & 1/4 & -1 & 1/2 \\ 1/4 & -1/2 & 1/2 & -2-2\mathrm{i} \end{bmatrix},$$

我们有

$$\mathscr{D}_1 = \{z \in \mathbb{C}: |z-2| \leqslant 1\}, \mathscr{D}_2 = \{z \in \mathbb{C}: |z-1-2\mathrm{i}| \leqslant 1/2\},$$

$$\mathscr{D}_3 = \{z \in \mathbb{C}: |z+1| \leqslant 5/4\}, \mathscr{D}_4 = \{z \in \mathbb{C}: |z+2+2\mathrm{i}| \leqslant 5/4\}.$$

如图 6.1 所示，$\mathscr{D}_1$ 与 $\mathscr{D}_2$ 为孤立的圆盘，因而各含有 A 的一个特征值，并集$G(A)$的第三个分枝为 $\mathscr{D}_3 \cup \mathscr{D}_4$，它含 A 的两个特征值.

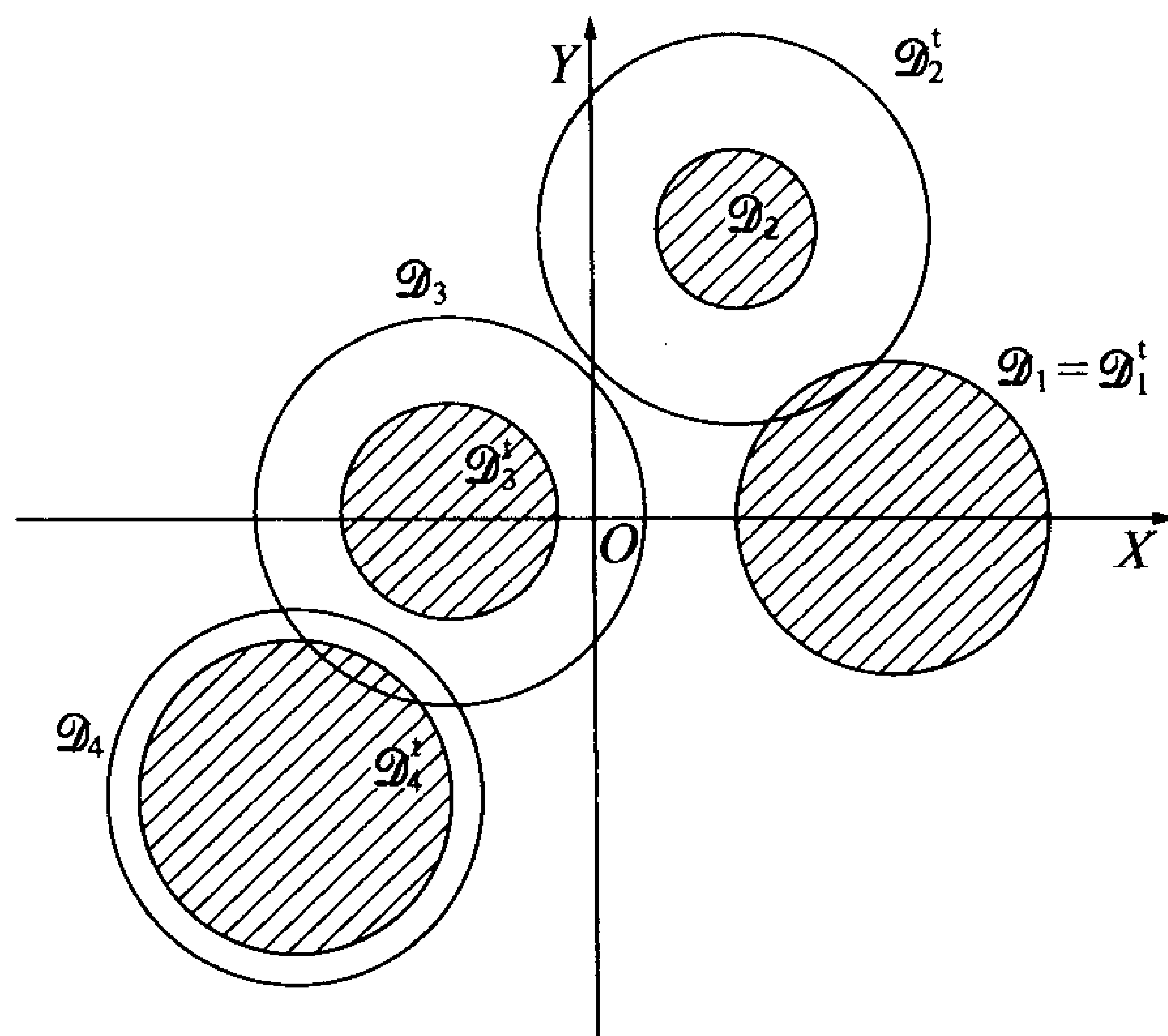


图 6.1　A 的 Gerschgorin 圆盘

如考虑 A 的转置矩阵 A^{T}（它与 A 有相同的特征值），则它的四个 Gerschgorin 圆盘为

$$\mathscr{D}_1^t = \mathscr{D}_1, \qquad \mathscr{D}_2^t = \{z \in \mathbb{C}: |z-1-2\mathrm{i}| \leqslant 5/4\},$$

$$\mathscr{D}_3^{\natural}=\{z\in\mathbb{C}:|z+1|\leqslant 3/4\},\qquad \mathscr{D}_4^{\natural}=\{z\in\mathbb{C}:|z+2+2\mathrm{i}|\leqslant 1\}.$$

如图 6.1 所示，$\mathscr{D}_3^{\natural}$与 $\mathscr{D}_4^{\natural}$为孤立的，因而含在 $\mathscr{D}_3\cup\mathscr{D}_4$ 内的两个特征值一个在 $\mathscr{D}_3^{\natural}$内，一个在 $\mathscr{D}_4^{\natural}$中. □

从前面讨论看出，定理 2 与定理 5 之间有着明显的关系. 应用定理 2 我们推出定理 5；反之，应用定理 5(第一部分)也能推出定理 2，这是因为定理 2 中条件(2)表明 0 属于 A 的排除区域$\mathbb{C}\backslash G(A)$，因而 $0\notin\sigma(A)$，亦即 $\det A\neq 0$.

任何矩阵的排除定理，至少像 Gerschgorin 圆盘定理以及在本小节内讨论的那些排除定理，皆可视为矩阵特征值的一种扰动定理(详见 6.3 节)，因为它们都给出当矩阵本身变化时，特征值变化的界. 这些都基于如下事实(它本身也是一种排除定理)，其中用到矩阵下界的概念(见第二章 2.3.2 小节).

定理 7　设 $A,B\in\boldsymbol{M}_n(\mathbb{C})$与 $\mu\in\mathbb{C}$. 则

$$\sigma(A+B)\subset\mathscr{D}_{A+B}\equiv\{z\in\mathbb{C}:\|((z-\mu)I-B)^{-1}(A-\mu I)\|\geqslant 1 \text{ 或 } \det((z-\mu)I-B)=0\},\tag{9}$$

其中，$\|\cdot\|$ 为 $\boldsymbol{M}_n(\mathbb{C})$上任一矩阵范数. 同时还有

$$\sigma(A+B)\subset\hat{\mathscr{D}}_{A+B}\equiv\{z\in\mathbb{C}:\mathrm{glb}((z-\mu)I-B)\leqslant\|A-\mu I\|\},\tag{10}$$

其中，$\mathrm{glb}(\cdot)$为关于$\mathbb{C}^n$ 上向量范数 $\|\cdot\|_\alpha$ 的矩阵下界，$\|\cdot\|$ 为 $\|\cdot\|_\alpha$ 导出的矩阵范数.

证明　设 $\lambda\in\sigma(A+B)$与 $\boldsymbol{x}\neq\boldsymbol{0}$ 为对应的特征向量，则有 $A\boldsymbol{x}=(\lambda I-B)\boldsymbol{x}$. 因此，

$$(A-\mu I)\boldsymbol{x}=((\lambda-\mu)I-B)\boldsymbol{x},\qquad\forall\mu\in\mathbb{C}.\tag{11}$$

若 $\lambda\in\sigma(\mu I+B)$，则 $\det((\lambda-\mu)I-B)=0$，因而有 $\lambda\in\mathscr{D}_{A+B}$. 若 $\lambda\notin\sigma(\mu I+B)$，则 $\det((\lambda-\mu)I-B)\neq 0$，因而由(11)式得出

$$\boldsymbol{x}=((\lambda-\mu)I-B)^{-1}(A-\mu I)\boldsymbol{x}.$$

按第二章 2.2.2 小节定理 3，对矩阵范数 $\|\cdot\|$，必存在与它相容的向量范数 $\|\cdot\|_\beta$. 于是，

$$\|\boldsymbol{x}\|_\beta\leqslant\|((\lambda-\mu)I-B)^{-1}(A-\mu I)\|\;\|\boldsymbol{x}\|_\beta,$$

这表明，$\lambda\in\mathscr{D}_{A+B}$，因为 $\|\boldsymbol{x}\|_\beta>0$.

对于$\mathbb{C}^n$ 上向量范数 $\|\cdot\|_\alpha$，按第二章习题 2.3.2 题 8，(11)式蕴涵

$$\mathrm{glb}((\lambda-\mu)I-B)\leqslant\|A-\mu I\|,$$

式中 $\mathrm{glb}(\cdot)$为关于 $\|\cdot\|_\alpha$ 的矩阵下界，$\|\cdot\|$ 为 $\|\cdot\|_\alpha$ 导出的矩阵范数，因而 $\lambda\in\hat{\mathscr{D}}_{A+B}$. □

上述定理提出矩阵 $C=A+B$ 的两个排除区域，一为$\mathbb{C}\backslash\mathscr{D}_{A+B}$，另一为$\mathbb{C}\backslash\hat{\mathscr{D}}_{A+B}$. 并且，当考虑不同向量或矩阵范数与 C 的不同分解 $A+B$ 时，我们可以得到多种的排除区域. 若将 B 视为原始矩阵，A 视为扰动，则(9)式与(10)式告诉我们，扰动后

的矩阵 $A+B$ 的特征值某种界可由原始矩阵 B 与扰动 A 的"信息"得出.

我们指出,定理 7 是比 Gerschgorin 圆盘定理更一般的排除定理,后者只是前者的特殊情形. 事实上,令 $\mu=0$,并取 B 为 $C=A+B$ 的对角部分,即 $B=\mathrm{diag}[c_{11}, c_{22},\cdots,c_{nn}]$; $\|\cdot\|=\|\cdot\|_\infty$,则(9)式这时变为

$$\sigma(C)\subset\mathscr{D}_c=\{z\in\mathbb{C}:\|(zI-B)^{-1}A\|_\infty\geqslant 1 \text{ 或 } \det(zI-B)=0\}$$
$$=\{z\in\mathbb{C}:\max_{1\leqslant i\leqslant n}\frac{\sum\limits_{j\neq i}|c_{ij}|}{|z-c_{ii}|}\geqslant 1 \text{ 或 } \prod_{i=1}^{n}|z-c_{ii}|=0\}.$$

容易看出,上式相当于

$$\sigma(C)\subset\mathscr{D}_c=G(C)\ .$$

习题 6.1.1

1. 试证:若 $A=(a_{ij})\in\boldsymbol{M}_n(\mathbb{R})(n\geqslant 2)$且

$$a_{ii}>R_i(A),\qquad i=1,2,\cdots,n,$$

则 $\det A>0$.

2. 设 $A=(a_{ij})\in\boldsymbol{M}_n(\mathbb{C})(n\geqslant 2)$为对角占优矩阵并有非负的对角元素. 证明: $\mathrm{Re}\lambda\geqslant 0$, $\forall\lambda\in\sigma(A)$. 此外,若 A 还是 Hermite 矩阵,则 A 为 Hermite 非负定的.

3. 设

$$A=\begin{bmatrix}1 & -1/2 & -1/2 & 0\\ -1/2 & 3/2 & \mathrm{i} & 0\\ 0 & -\mathrm{i}/2 & 5 & \mathrm{i}/2\\ -1 & 0 & 0 & 5\mathrm{i}\end{bmatrix}.$$

试尽力地估计 A 的特征值位置.

4. 证明:若 $A=(a_{ij})\in\boldsymbol{M}_n(\mathbb{C})(n\geqslant 2)$为严格对角占优矩阵,则矩阵 $I-H^{-1}A$ 的所有特征值的模小于 1,其中 $H=\mathrm{diag}[a_{11},a_{22},\cdots,a_{nn}]$.

5. 设 $A=(a_{ij})\in\boldsymbol{M}_n(\mathbb{C})$有特征值 $\lambda_1,\cdots,\lambda_n$,且 $\|\cdot\|_F$ 表示 $\boldsymbol{M}_n(\mathbb{C})$上的 Frobenius 范数. 试证:

$$\sum_{i=1}^{n}|\lambda_i|^2\leqslant\|A\|_F^2,$$
$$\sum_{i=1}^{n}|\mathrm{Re}\lambda_i|^2\leqslant\|\mathrm{Re}A\|_F^2,$$
$$\sum_{i=1}^{n}|\mathrm{Im}\lambda_i|^2\leqslant\|\mathrm{Im}A\|_F^2.$$

并且,上述任一等式成立当且仅当 A 为正规矩阵. 这里, $\mathrm{Re}A$ 与 $\mathrm{Im}A$ 分别表示矩阵 A 的实部与虚部.

6. 设 $A=(a_{ij})\in\boldsymbol{M}_n(\mathbb{C})$, $B=\mathrm{Re}A=(b_{ij})$(矩阵 A 的实部),与 $C=\mathrm{Im}A=(c_{ij})$. 证明: $|\lambda|\leqslant n\max\limits_{i,j}|a_{i,j}|$, $|\mathrm{Re}\lambda|\leqslant n\max\limits_{i,j}|b_{ij}|$, $|\mathrm{Im}\lambda|\leqslant n\max\limits_{i,j}|c_{i,j}|$, $\forall\lambda\in\sigma(A)$.
若 $A+A^{\mathrm{T}}\in\boldsymbol{M}_n(\mathbb{R})$,则上述第三个不等式可以改善为

$$|\operatorname{Im}\lambda| \leqslant \max_{i,j}|c_{i,j}|\sqrt{\frac{n(n-1)}{2}}, \qquad \forall \lambda \in \sigma(A).$$

7. 设 $A=(a_{ij})\in \boldsymbol{M}_n(\mathbb{R})$，$\tau=\frac{1}{2}\max\limits_{i,j}|a_{ij}-a_{ji}|$. 试证：

$$|\operatorname{Im}\lambda| \leqslant \tau\sqrt{n(n-1)/2}, \qquad \forall \lambda \in \sigma(A).$$

8. 设 $A=(a_{ij})\in \boldsymbol{M}_n(\mathbb{C})$ 有特征值 $\lambda_1,\cdots,\lambda_n$ 与奇异值 $\alpha_1\geqslant\cdots\geqslant\alpha_n$. 试证：$\alpha_n\leqslant|\lambda_i|\leqslant\alpha_1$，$i=1,\cdots,n$.

9. 设 $A=(a_{ij})\in \boldsymbol{M}_n(\mathbb{C})$ 有特征值 $\lambda_1,\cdots,\lambda_n$ 与奇异值 $\alpha_1\geqslant\cdots\geqslant\alpha_n$. 试证：

$$\prod_{i=1}^{k}\alpha_{n-i+1} \leqslant \prod_{s=1}^{k}|\lambda_{i_s}| \leqslant \prod_{i=1}^{k}\alpha_i, \qquad \forall (i_1,\cdots,i_k)\in Q_{k,n}, k=1,2,\cdots,n.$$

10. 已知三阶复矩阵

$$A=\begin{bmatrix} 7+3\mathrm{i} & -4-6\mathrm{i} & -4 \\ -1-6\mathrm{i} & 7 & -2-6\mathrm{i} \\ 2 & 4-6\mathrm{i} & 13-3\mathrm{i} \end{bmatrix}$$

有 $\sigma(A)=\{9,9\pm 9\mathrm{i}\}$. 试用题 5，题 6，题 7，以及定理 5，给出 A 的特征值及其实部与虚部的估计.

11. 试证：若 $A=(a_{ij})\in \boldsymbol{M}_n(\mathbb{C})$ $(n\geqslant 2)$ 严格对角占优，则

$$|\det A| \geqslant \prod_{i=1}^{n}[|a_{ii}|-R_i(A)].$$

6.1.2　不可约矩阵的情形

先引入不可约矩阵的定义与有关特征.

定义 1　$A=(a_{ij})\in \boldsymbol{M}_n(\mathbb{C})$ 称为**可约的**(reducible)，假如它是一阶零矩阵或者对 $n\geqslant 2$ 有置换矩阵 P 使得

$$P^{\mathrm{T}}AP=\begin{bmatrix} B & C \\ O & D \end{bmatrix}, \tag{1}$$

式中，B 与 D 为阶数 $\geqslant 1$ 的方阵. 不是可约的矩阵 $A=(a_{ij})\in \boldsymbol{M}_n(\mathbb{C})$ 叫作**不可约的**(irreducible).

按这个定义，各阶零矩阵为可约的，而 $\boldsymbol{M}_n(\mathbb{C})$ 内正矩阵必为不可约的. 当 $A\in \boldsymbol{M}_n(\mathbb{C})$ $(n\geqslant 2)$ 可约时，线性方程组 $A\boldsymbol{x}=\boldsymbol{b}$ 的方程与未知量的次序作同步变换之后，可化为两个较为低阶的方程组. 实际上，按(1)式，$P^{\mathrm{T}}APP^{\mathrm{T}}\boldsymbol{x}=P^{\mathrm{T}}\boldsymbol{b}$ 有形式

$$\begin{bmatrix} B & C \\ O & D \end{bmatrix}\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{z} \end{bmatrix}=\begin{bmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \end{bmatrix},$$

这里，$\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{z} \end{bmatrix}=P^{\mathrm{T}}\boldsymbol{x}$ 与 $\begin{bmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \end{bmatrix}=P^{\mathrm{T}}\boldsymbol{b}$，即有

$$B\boldsymbol{y}+C\boldsymbol{z}=\boldsymbol{b}_1,$$
$$D\boldsymbol{z}=\boldsymbol{b}_2.$$

下面定理 2 与定理 4 给出不可约矩阵的两个特征.

定理 2　设 $A=(a_{ij})\in \boldsymbol{M}_n(\mathbb{C})$，$n\geqslant 2$. 则 A 不可约当且仅当对集合 $\mathcal{N}=\{1,2,\cdots,n\}$ 的任意劈分 $\mathcal{S}$ 与 $\mathcal{T}$(即 $\mathcal{S}$ 与 $\mathcal{T}$ 不空不交，且 $\mathcal{S}\cup\mathcal{T}=\mathcal{N}$)，$A$ 有元素 $a_{ij}\neq 0$ 使得 $i\in\mathcal{S}$ 与 $j\in\mathcal{T}$. 或等价地，A 可约当且仅当有 $\mathcal{N}$ 的一种劈分 $\mathcal{S}$ 与 $\mathcal{T}$，使得 $a_{ij}=0$，$\forall i\in\mathcal{S}$ 与 $\forall_j\in\mathcal{T}$.

证明　若有 $\mathcal{N}$ 的劈分 $\mathcal{S}$ 与 $\mathcal{T}$ 使得 $a_{ij}=0$，$\forall i\in\mathcal{S}$ 与 $\forall j\in\mathcal{T}$，则显然有置换矩阵 P，使得 $P^{\mathrm{T}}AP$ 有形式(1)，其中，B 与 D 的行序分别对应于 $\mathcal{T}$ 与 $\mathcal{S}$ 的元素，因而 A 为可约矩阵. 反之，若 A 可约，按定义，则(1)式对某个置换矩阵 P 成立. 设方阵 D 对应 A 的行序的集合为 $\mathcal{S}=\{i_1,\cdots,i_s\}$，$1\leqslant s<n$，且令 $\mathcal{T}=\mathcal{N}-\mathcal{S}$. 则 $\mathcal{S}$ 与 $\mathcal{T}$ 为 $\mathcal{N}$ 的一种劈分，且对任意 $i\in\mathcal{S}$ 与 $j\in\mathcal{T}$ 都有 $a_{ij}=0$. □

应用定义 1 与定理 2 可以判定矩阵的不可约性，但更常用的一种方法为考察矩阵有向图的连接性.

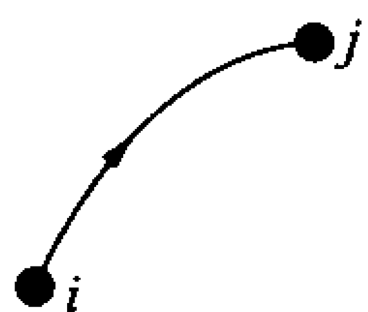


图 6.2　有向弧线

一个**有向图** Γ 可以视为**结点**(node)以及连接结点**有向弧线**的集合. 在本书内主要讨论矩阵 $A\in\boldsymbol{M}_n(\mathbb{C})$ 的有向图，记为 $\Gamma(A)$，其构造如下. 在平面上取 n 个不同点标以号码 $1,2,\cdots,n$，称之为 $\Gamma(A)$ 的**结点**(或**顶点**)，对 A 的每一个非零元素 a_{ij}，用一条从结点 i 指向结点 j 的有向弧线来连接结点 i 与 j(图 6.2). 例如，

$$A=\begin{bmatrix}0 & -2 & -1\\ 3 & 0 & 0\\ -1 & 0 & 0\end{bmatrix}\quad 与\quad B=\begin{bmatrix}2 & 1\\ 0 & -1\end{bmatrix}$$

的有向图 $\Gamma(A)$ 与 $\Gamma(B)$ 如图 6.3 所示.

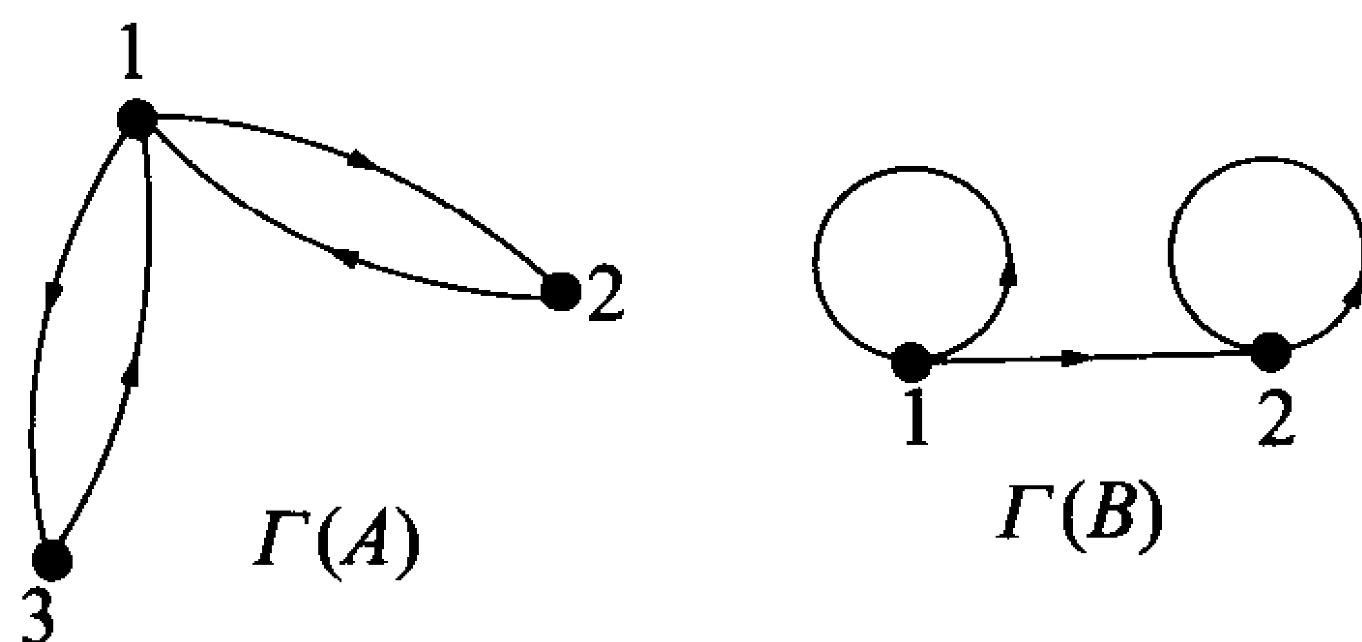


图 6.3　有向图 $\Gamma(A)$ 与 $\Gamma(B)$

容易看出，对 $A\in\boldsymbol{M}_n(\mathbb{C})$ 施以行与列同步变换只是改变 $\Gamma(A)$ 中结点的排列次序.

定义 3　有向图 Γ 叫作**强连接的**，假如对 Γ 的任何一对有序结点 (i,j)，$i\neq j$，都有一条由有向弧线组成的**有向路**(directed path)

$$\widehat{il_1},\widehat{l_1l_2},\cdots,\widehat{l_{r-1}j}\tag{2}$$

连接结点 i 与 j.

例如,图 6.3 中 $\Gamma(A)$为强连接的,而 $\Gamma(B)$不是.

定理 4 设 $A=(a_{ij})\in \boldsymbol{M}_n(\mathbb{C}),n\geqslant 2$. 则 A 不可约当且仅当 A 的有向图 $\Gamma(A)$ 强连接.

证明 设 A 不可约且令 $\mathcal{N}=\{1,2,\cdots,n\}$, $\mathcal{X}_i=\{j\in\mathcal{N}: j\neq i$ 且从 i 有一有向路到 $j\}$, $\forall i\in\mathcal{N}$. 要证 $\Gamma(A)$强连接只要证 $\mathcal{X}_i=\mathcal{N}-\{i\}$, $\forall i\in\mathcal{N}$. 取 $\mathcal{S}_1=\{i\}$与 $\mathcal{T}_1=\mathcal{N}-\mathcal{S}_1$,则它们为 $\mathcal{N}$ 的劈分,按定理 2,有 $l_1\in\mathcal{T}_1$ 使得 $a_{il_1}\neq 0$,因而 $l_1\in\mathcal{X}_i$. 若 $n=2$,则 $\mathcal{X}_i=\{1,2\}-\{i\}$,因而 $\Gamma(A)$强连接. 若 $n\geqslant 3$,令 $\mathcal{S}_2=\{i,l_1\}$与 $\mathcal{T}_2=\mathcal{N}-\mathcal{S}_2$,则它们为 $\mathcal{N}$ 的一种劈分,按定理 2,有 $l_2\in\mathcal{T}_2$ 使得 $a_{il_2}\neq 0$ 或者 $a_{l_1l_2}\neq 0$,因而 $l_2\in\mathcal{X}_i$. 继续这个过程,可证 $\mathcal{X}_i=\mathcal{N}-\{i\}$, $\forall i\in\mathcal{N}$. 反之,设 $\Gamma(A)$强连接且 $\mathcal{S}$与 $\mathcal{T}$ 为 $\mathcal{N}$ 的任意劈分. 任取 $i\in\mathcal{S}$与 $j\in\mathcal{T}$,按 $\Gamma(A)$强连接性,或有 $a_{ij}\neq 0$ 或有 $l_1,l_2,\cdots,l_r\in\mathcal{N}$使得 $a_{il_1}a_{l_1l_2}\cdots a_{l_rj}\neq 0$. 对后一种情形,$a_{il_1},\cdots,a_{l_rj}$ 之中至少有一个元素,它的第一个下标属于 $\mathcal{S}$,第二个下标属于 $\mathcal{T}$. 因此,总有 $a_{pq}\neq 0$ 使 $p\in\mathcal{S}$与 $q\in\mathcal{T}$. 这时定理 2 断定 A 是不可约的. □

作为定理 2 与定理 4 的直接推论,我们有

推论 5 设 $A=(a_{ij})\in \boldsymbol{M}_n(\mathbb{C})$不可约. 则 $A^{\mathrm{T}}=(a_{ji})$也不可约,且对 $n\geqslant 2$,有 $R_i(A)>0$ 与 $C_i(A)>0$, $i=1,\cdots,n$.

关于不可约矩阵的非奇异性定理与对应的排除定理比起一般复方阵的相应结果来往往有一些更好的结论. 为此,先考虑不可约对角占优矩阵的定义.

定义 6 若 $A\in \boldsymbol{M}_n(\mathbb{C})\,(n\geqslant 2)$不可约,对角占优,且 6.1.1 小节(1)式中至少有一个严格不等式成立,则 A 称为**不可约对角占优**矩阵. 类似地,矩阵 $A\in \boldsymbol{M}_n(\mathbb{C})$ 称为不可约(列)对角占优的,假若 A^{T} 为不可约对角占优矩阵.

下面是关于不可约矩阵的一种排除定理.

定理 7(Taussky) 设 $A=(a_{ij})\in \boldsymbol{M}_n(\mathbb{C})\,(n\geqslant 2)$不可约. 则 A 的任何一个特征值均不能落在 6.1.1 小节(6)式中诸圆盘 $\mathcal{D}_i$ 之并集 $G(A)$的边界上,除非它同时落在每个圆盘 $\mathcal{D}_i$ 的边界上.

证明 设 $\omega\in\sigma(A)$落在 $G(A)$的边界上,且 $\boldsymbol{x}=(x_1,\cdots,x_n)^{\mathrm{T}}\neq\boldsymbol{0}$ 为对应的特征向量. 令 $|x_r|=\max\limits_i|x_i|$,则 $x_r\neq 0$. 这时,方程组 $A\boldsymbol{x}=\omega\boldsymbol{x}$ 的第 r 个方程蕴涵

$$|\omega-a_{rr}||x_r|\leqslant\sum_{j\neq r}|a_{rj}||x_j|\leqslant\sum_{j\neq r}|a_{rj}||x_r|.$$

按假定 ω 为 $G(A)$的边界点,故有 $|\omega-a_{rr}|=R_r(A)$,因而(3)式变为等式,且有

$$|\omega-a_{rr}|=\sum_{j\neq r}|a_{rj}||x_j|/|x_r|=R_r(A).$$

由此推出,凡满足 $a_{rl}\neq 0\,(l\neq r)$的 $l\in\mathcal{N}=\{1,\cdots,n\}$,都有 $|x_l|=|x_r|$. 因为 A 不可约,推论 5 断定至少有一个这样的 $l\in\mathcal{N}$,取 $l(\neq r)$代替 r 并重复前面的推证,可得

$$|\omega - a_{ll}| = \sum_{j\neq l} |a_{lj}|\,|x_j| / |x_l| = R_l(A).$$

于是,凡满足 $a_{lp}\neq 0(p\neq l)$的 $p\in\mathcal{N}$都有$|x_p|=|x_l|=|x_r|$. 按定理 4,对任意 $j\in\mathcal{N}, j\neq r$,都有 A 的非零元素 $a_{rl}, a_{lp}, \cdots, a_{qj}$,它们保证上述推证过程经过有限步以后便得$|x_j|=|x_r|$. 因此,

$$|\omega - a_{jj}| = R_j(A), \qquad \forall j\in\mathcal{N},$$ □

即 ω 为每个圆盘 $\mathcal{D}_j$ 的边界点.

与此排除定理相对应的是如下关于不可约矩阵的非奇异性定理.

定理 8(Taussky)　若 $A=(a_{ij})\in \boldsymbol{M}_n(\mathbb{C})(n\geqslant 2)$不可约对角占优,则 $\det A\neq 0$,且所有 $a_{ii}\neq 0$.

证明　从 6.1.1 小节(1)式与推论 5 直接推出 $a_{ii}\neq 0, i=1,\cdots,n$. 倘若 $\det A=0$,则 $A\boldsymbol{x}=\boldsymbol{0}$ 有非零解 $\boldsymbol{x}=(x_1,\cdots,x_n)^{\mathrm{T}}$. 令$|x_r|=\max\limits_i|x_i|(\neq 0)$. 从定理 7 证明看出,$|a_{rr}|\leqslant R_r(A)$. 但由于 A 满足 6.1.1 小节(1)式,故有$|a_{rr}|=R_r(A)$,即 A 的零特征值位于圆盘 $\mathcal{D}_r$ 的边界上. 按照定理 7 的证明,$\omega=0$ 必定位于每个圆盘 $\mathcal{D}_j$ 的边界上:$|a_{jj}|=R_j(A), j=1,\cdots,n$,此与定理假定矛盾. 因此,$\det A\neq 0$. □

例 9　下列 n 阶($n\geqslant 2$)三对角矩阵

$$A=\begin{bmatrix} 2 & -1 & & & 0\\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1\\ 0 & & & -1 & 2\end{bmatrix}$$

为不可约对角占优矩阵,因而 $\det A\neq 0$. 更一般地,满足条件(令 $a_1=0$):

$$|b_1|\geqslant|c_1|, |b_n|>|a_n|, |b_i|\geqslant|a_i|+|c_i|, c_i\neq 0, i=1,2,\cdots,n-1$$

且阶数 $n\geqslant 2$ 的三对角矩阵

$$B=\begin{bmatrix} b_1 & c_1 & & & 0\\ a_2 & b_2 & c_2 & & \\ & \ddots & \ddots & \ddots & \\ & & a_{n-1} & b_{n-1} & c_{n-1}\\ 0 & & & a_n & b_n\end{bmatrix}$$

也是非奇异的,并且,B 的所有前主子阵也是非奇异的. 事实上,若 $a_2a_3\cdots a_n\neq 0$,则从 B 的有向图看出,B 是不可约的,于是按题设条件,B 为不可约对角占优矩阵,因而 $\det B\neq 0$. 若 $a_2a_3\cdots a_n=0$,则 $\det B$ 等于若干个不可约对角占优矩阵的行列式之积,因而 $\det B\neq 0$. 类似可证余下的. □

由于 A 不可约蕴涵它的任何正对角相似矩阵 $D^{-1}AD$(D 为正对角矩阵)的不可约性,故从定理 8 立即推出

推论 10　设 $A=(a_{ij})\in \boldsymbol{M}_n(\mathbb{C})(n\geqslant 2)$不可约. 如果存在正对角矩阵 D 使得 $D^{-1}AD$ 对角占优，且$|a_{ii}|\geqslant R_i(D^{-1}AD)$，$i=1,2,\cdots,n$ 中至少有一个严格不等式，则 $\det A\neq 0$，且所有 $a_{ii}\neq 0$.

习题 6.1.2

1. 设$A=(a_{ij})\in \boldsymbol{M}_n(\mathbb{C})$不可约，且至少有一个 i 使得$\sum_{j=1}^{n}|a_{ij}|<\|A\|_\infty$. 试证：$\rho(A)<\|A\|_\infty$.

2. 画出下列矩阵的有向图，并判别它们的不可约性：

$$A=\begin{bmatrix}0&1&0&1\\1&0&1&0\\0&1&0&1\\1&0&0&1\end{bmatrix},\quad B=\begin{bmatrix}0&0&1&1\\0&0&1&1\\1&1&0&0\\1&1&0&0\end{bmatrix},\quad C=\begin{bmatrix}1&1&0&0\\0&2&1&0\\0&0&3&1\\1&0&0&4\end{bmatrix}.$$

3. 设 $A=(a_{ij})\in \boldsymbol{M}_n(\mathbb{C})$不可约且对角占优，又设所有 $a_{ii}\geqslant 0$ 与 $a_{ij}\leqslant 0(i\neq j)$. 证明：当且仅当

$$\sum_{j=1}^{n}a_{ij}=0,\qquad i=1,2,\cdots,n$$

时，$0\in\sigma(A)$，即有 $\det A=0$.

4. 设 $A=(a_{ij})\in \boldsymbol{M}_n(\mathbb{C})(n\geqslant 2)$不可约且对角占优. 证明：$\operatorname{rank}A\geqslant n-1$.

5. 证明：设 $A=(a_{ij})\in \boldsymbol{M}_n(\mathbb{C})(n\geqslant 2)$对角占优，且 $J=\{j\in\mathcal{N}:|a_{jj}|>R_j(A)\}$不空. 若对任意 $i\notin J$，存在 $\Gamma(A)$中一条有向路从 i 到 $j\in J$，则 $\det A\neq 0$.

6. 矩阵 $A=(a_{ij})\in \boldsymbol{M}_n(\mathbb{C})(n\geqslant 2)$称为**下半严格对角占优的**，假如它对角占优，并且，

$$|a_{ii}|>\sum_{j=1}^{i-1}|a_{ij}|,\qquad i=2,3,\cdots,n.$$

矩阵 $A=(a_{ij})\in \boldsymbol{M}_n(\mathbb{C})(n\geqslant 2)$称为**半严格对角占优的**，假如有置换矩阵 P 使得 $P^{\mathrm{T}}AP$ 为下半严格对角占优的. 试证：当 $n\geqslant 2$ 时，$A\in \boldsymbol{M}_n(\mathbb{C})$为不可约对角占优当且仅当 A 为半严格对角占优与不可约的矩阵.

7. 证明：半严格对角占优的复矩阵为非奇异的.

8. 应用题 1 结果证明：若多项式

$$p(z)=z^n+a_{n-1}z^{n-1}+\cdots+a_1z+a_0,\qquad a_0\neq 0$$

不满足条件

$$|a_0|=|a_1|+1=|a_2|+1=\cdots=|a_{n-1}|+1,$$

则 $p(z)$的任意零点 $\tilde{z}$ 有界如下：

$$|\tilde{z}|<\max\{|a_0|,|a_1|+1,|a_2|+1,\cdots,|a_{n-1}|+1\}.$$

6.2　对角占优矩阵的推广及其相应的排除定理

6.2.1　Brauer 定理与 Ostrowski 定理

6.1 节中讨论的矩阵非奇异性定理与相应的排除定理已经有许多变种与推

广,本小节介绍的 Brauer 定理与 Ostrowski 定理就是其中重要的部分.

定理 1(Brauer)　若 $A=(a_{ij})\in \boldsymbol{M}_n(\mathbb{C})(n\geqslant 2)$满足

$$|a_{ii}||a_{jj}|>R_i(A)R_j(A),\qquad i,j=1,\cdots,n;i\neq j,\tag{1}$$

则 $\det A\neq 0$.

证明　设条件(1)成立但 $\det A=0$. 则 $A\boldsymbol{x}=\boldsymbol{0}$ 有非零解 $\boldsymbol{x}=(x_1,\cdots,x_n)^{\mathrm{T}}$. 令 $r\neq s$使得$|x_r|\geqslant|x_s|\geqslant|x_i|$, $\forall\, i\in\mathcal{N}, i\neq r$. 此时有 $x_s\neq 0$,否则对所有 $i\neq r, x_i=0$,于是由 $x_r\neq 0$ 与 $A\boldsymbol{x}=\boldsymbol{0}$ 的第 r 个方程得出 $a_{rr}x_r=0$,因而 $a_{rr}=0$,与(1)式抵触. 从 $A\boldsymbol{x}=\boldsymbol{0}$ 的第 r 与 s 个方程推得

$$|a_{rr}||x_r|\leqslant\sum_{j\neq r}|a_{rj}||x_j|\leqslant R_r(A)|x_s|,$$

$$|a_{ss}||x_s|\leqslant\sum_{j\neq s}|a_{sj}||x_j|\leqslant R_s(A)|x_r|.$$

相乘之后约去公因子便有$|a_{rr}||a_{ss}|\leqslant R_r(A)R_s(A)$,与(1)式矛盾. 因此,$\det A\neq 0$.

□

显然 6.1.1 小节中条件(2)蕴涵这里的条件(1),因而上述定理确为 Lévy-Desplanques 定理的一种推广. 类似地,如用 $C_i(A)$代替(1)式中 $R_i(A), i=1,2,\cdots,n$,或者用 $R_i(D^{-1}AD)$(D 为正对角矩阵)代替(1)式中 $R_i(A), i=1,2,\cdots,n$,则也能得到相应的推广.

对应于定理 1,我们有下列排除定理,它是 Gerschgorin 圆盘定理的一种改进.

定理 2(Brauer 卵形定理)　设 $A=(a_{ij})\in \boldsymbol{M}_n(\mathbb{C})(n\geqslant 2)$. 则 A 的谱$\sigma(A)$包含在复平面上 $n(n-1)/2$ 个 Cassini 卵形区域 $\mathscr{D}_{i,j}$的并集内,其中,

$$\mathscr{D}_{i,j}=\{z\in\mathbb{C}:|z-a_{ii}||z-a_{jj}|\leqslant R_i(A)R_j(A)\}, i,j=1,\cdots,n;i\neq j.\tag{2}$$

此外,若用 $\hat{\mathscr{D}}_{i,j}$表示卵形 $\mathscr{D}_{i,j}$包含焦点 a_{ii}的单连通部分,且 $\mathscr{H}_i=\bigcup_{j\neq i}\hat{\mathscr{D}}_{i,j}, i=1,2,\cdots,n$,则只要 f 个 $\mathscr{H}_i$ 组成的区域 $\mathscr{R}$ 与其余的 $\mathscr{H}_i$ 不交,$\mathscr{R}$ 便恰含有 A 的 f 个特征值.

本定理的证明留给读者(见习题 1).

现在证明 Ostrowski 的两个定理,它们都属于矩阵非奇异性定理. 先引进一些记号. 对 $A=(a_{ij})\in \boldsymbol{M}_n(\mathbb{C})(n\geqslant 2)$与 $s>0$,记

$$\begin{aligned}R_{i,s}(A)&=\Big(\sum_{j\neq i}|a_{ij}|^s\Big)^{1/s},\qquad C_{i,s}(A)=\Big(\sum_{j\neq i}|a_{ji}|^s\Big)^{1/s},\\ R_{i,\infty}(A)&=\max_{j\neq i}|a_{ij}|,\qquad C_{i,\infty}(A)=\max_{j\neq i}|a_{ji}|.\end{aligned}\tag{3}$$

回顾一下,若 $p\geqslant 1$ 与 $q\geqslant 1$ 满足 $p^{-1}+q^{-1}=1$,则 p 与 q 为一对共轭指数. 如 2 与 2,1 与∞各成一对共轭指数.

定理 3(Ostrowski)　设 $\alpha\in(0,1)$, p 与 q 为一对共轭指数. 若 $A=(a_{ij})\in \boldsymbol{M}_n(\mathbb{C})(n\geqslant 2)$满足

$$|a_{ii}|>R_{i,\alpha p}^{\alpha}(A)C_{i,(1-\alpha)q}^{1-\alpha}(A),\qquad i=1,\cdots,n,\tag{4}$$

则 $\det A \neq 0$.

证明　不失普遍性假定 $R_{i,\alpha p}(A) \neq 0, \forall i \in \mathcal{N}$.(因为如有 $i \in \mathcal{N}$ 使得 $R_{i,\alpha p}(A)=0$,则 $a_{ij}=0, \forall j \neq i$,于是问题基本上归结为 $n-1$ 阶的情形)用反证法. 假如 $\det A=0$,则有 $\boldsymbol{x}=(x_1,\cdots,x_n)^{\mathrm{T}} \neq \boldsymbol{0}$ 使得 $A\boldsymbol{x}=\boldsymbol{0}$. 因此,

$$|a_{ii}||x_i| \leqslant \sum_{j \neq i} |a_{ij}||x_j|, \qquad i=1,\cdots,n.$$

应用(4)式与 Hölder 不等式,从上式推出,对任意 $i \in \mathcal{N}$,

$$\begin{aligned} R_{i,\alpha p}^{\alpha}(A) C_{i,(1-\alpha)q}^{1-\alpha}(A)|x_i| &\leqslant \sum_{j \neq i} |a_{ij}|^{\alpha} |a_{ij}|^{1-\alpha} |x_j| \\ &\leqslant R_{i,\alpha p}^{\alpha}(A) \Big(\sum_{j \neq i} |a_{ij}|^{(1-\alpha)q} |x_j|^q \Big)^{1/q}, \end{aligned}$$

上式前面 n 个不等式中至少有一严格不等式成立,因为至少有一个 $x_i \neq 0$. 由于 $R_{i,\alpha p}(A) \neq 0$,故有

$$C_{i,(1-\alpha)q}^{(1-\alpha)q}(A)|x_i|^q \leqslant \sum_{j \neq i} |a_{ij}|^{(1-\alpha)q} |x_j|^q, \qquad i=1,\cdots,n.$$

相加之,我们有严格不等式

$$\begin{aligned} \sum_{i=1}^{n} C_{i,(1-\alpha)q}^{(1-\alpha)q}(A)|x_i|^q &< \sum_{i=1}^{n} \sum_{j \neq i} |a_{ij}|^{(1-\alpha)q} |x_j|^q \\ &= \sum_{i=1}^{n} \sum_{j \neq i} |a_{ji}|^{(1-\alpha)q} |x_i|^q = \sum_{i=1}^{n} C_{i,(1-\alpha)q}^{(1-\alpha)q}(A)|x_i|^q, \end{aligned}$$

这是不可能的. 因此,$\det A \neq 0$. □

我们指出,应用定理 3 可以判断某些不是严格对角占优矩阵的非奇异性. 例如,

$$A = \begin{bmatrix} 3 & 4 \\ -1 & 3 \end{bmatrix}$$

不是行或列严格对角占优的,但对 $\alpha=1/2$ 与 $p=q=2$,A 满足条件(4),因而 $\det A \neq 0$. 此外,定理 3 含有两种重要的特殊情形. 其一,取 $p=1/\alpha$(此时,$q=1/(1-\alpha)$),则(4)式变为

$$|a_{ii}| > R_i^{\alpha}(A) C_i^{1-\alpha}(A), \qquad i=1,\cdots,n. \tag{5}$$

而(行)严格对角占优与(列)严格对角占优条件分别可视为(5)式中取 $\alpha=1$ 与 $\alpha=0$ 的极限情形. 其二,取 $p=1$(此时,$q=\infty$)或取 $q=1$(此时,$q=\infty$),条件(4)变为

$$|a_{ii}| > R_{i,\alpha}^{\alpha}(A) C_{i,\infty}^{1-\alpha}(A), \qquad i=1,\cdots,n, \tag{6}$$

或者

$$|a_{ii}| > R_{i,\infty}^{\alpha}(A) C_{i,(1-\alpha)}^{1-\alpha}(A), \qquad i=1,\cdots,n. \tag{7}$$

定理 3 对应如下的排除定理.

定理 4　设 α, p 与 q 如同定理 3. 则 $A=(a_{ij}) \in \boldsymbol{M}_n(\mathbb{C})$ 的谱 $\sigma(A)$ 包含在复平面内下列 n 个圆盘的并集内:

$$\{z \in \mathbb{C}: |z - a_{ii}| \leqslant R_{i,\alpha p}^{\alpha}(A) C_{i,(1-\alpha)q}^{1-\alpha}(A)\}, \qquad i = 1, \cdots, n. \tag{8}$$

对于不可约矩阵情形，上述定理 3 与定理 4 也有更强的相应结果(见习题 2).

现在讨论 Ostrowski 的另一个定理，它类似于 Lévy-Desplanques 定理的 Brauer 推广(定理 1).

定理 5(Ostrowski) 设 α, p 与 q 如同定理 3. 若 $A = (a_{ij}) \in \boldsymbol{M}_n(\mathbb{C})(n \geqslant 2)$ 满足

$$|a_{ii}||a_{jj}| > f_i(A) f_j(A), \qquad i, j = 1, \cdots, n; i \neq j, \tag{9}$$

式中，

$$f_i(A) = R_{i,\alpha p}^{\alpha}(A) C_{i,(1-\alpha)q}^{1-\alpha}(A), \qquad i = 1, \cdots, n, \tag{10}$$

则 $\det A \neq 0$.

证明 令 $s_i = f_i(A)/|a_{ii}|$, $\forall i \in \mathcal{N}$, 并将它们排列为

$$s_{k_1} \geqslant s_{k_2} \geqslant \cdots \geqslant s_{k_n}.$$

此时，条件(9)等价于 $s_i s_j < 1, i, j = 1, \cdots, n; i \neq j$. 特别地有 $s_{k_1} s_{k_2} < 1$, 因而 $s_{k_2} < 1$. 假如 $s_{k_2} = 0$, 则 $s_{k_i} = 0$, $\forall i \geqslant 3$, 因而 $\det A = a_{11} \cdots a_{nn} \neq 0$. 假如 $0 < s_{k_2} < 1$, 令 $Q = (s_{k_1}/s_{k_2})^{1/2} (\geqslant 1)$, 并用 B 表示 Q 乘上 A 的第 k_1 行与第 k_1 列后得到的矩阵. 由于 $\det A = \det B / Q^2$, 故只要证明 $\det B \neq 0$ 即可. 对于 B, 约定采用 A 的类似的记号，不过都标以“′”以示区别. 此时，

$$|a'_{k_1 k_1}| = Q^2 |a_{k_1 k_1}|, \qquad R'_{k_1, \alpha p} = Q R_{k_1, \alpha p}, \qquad C'_{k_1, (1-\alpha) q} = Q C_{k_1, (1-\alpha) q}.$$

于是，$s'_{k_1} = s_{k_1}/Q = (s_{k_1} s_{k_2})^{1/2} < 1$. 对于 $i \geqslant 2$, $s'_{k_i} \leqslant Q s_{k_i} \leqslant Q s_{k_2} = (s_{k_1} s_{k_2})^{1/2} < 1$. 依照定理 3, $\det B \neq 0$, 因而 $\det A \neq 0$. □

定理 5 相应的排除定理如下.

定理 6 设 α, p 与 q 如同定理 3, 且 $f_i(A)$ 由(10)式确定. 则任意 $A = (a_{ij}) \in \boldsymbol{M}_n(\mathbb{C})(n \geqslant 2)$ 的谱 $\sigma(A)$ 包含在下列 $n(n-1)/2$ 个 Cassini 卵形的并集内：

$$\{z \in \mathbb{C}: |z - a_{ii}||z - a_{jj}| \leqslant f_i(A) f_j(A)\}, i, j = 1, \cdots, n; i \neq j. \tag{11}$$

习题 6.2.1

1. 证明定理 2.
2. 对于不可约矩阵 $A \in \boldsymbol{M}_n(\mathbb{C})$ 的情形，试写出并证明定理 3 与定理 4 的相应的结果.
3. 写出并证明与定理 1 相对应的不可约矩阵的非奇异性定理.

6.2.2 Shemesh 定理与 Brualdi 定理

受 6.1.1 小节定理 2 与 6.2.1 小节定理 1 启发，我们自然地提出问题：当 $k (\leqslant n-1)$ 为非负整数，且 $\mathcal{S}$ 取遍 $\mathcal{N} = \{1, 2 \cdots, n\}$ 的含有 $k+1$ 个不同元素的任意子集时，满足条件

$$\prod_{i \in \mathcal{S}} |a_{ii}| > \prod_{i \in \mathcal{S}} R_i(A), \qquad \forall \mathcal{S} \tag{1}$$

的矩阵 $A=(a_{ij})\in \boldsymbol{M}_n(\mathbb{C})(n\geqslant 2)$一定是非奇异的吗？对于 $k=0$ 与 $k=1$ 情形，(1)式分别变为 6.1.1 小节(2)式与 6.2.1 小节(1)式，因而此问题的答案是肯定的. 但对 $k\geqslant 2$ 情形，此问题的答案是否定的. 例如，对如下 n 阶($n\geqslant 3$)奇异矩阵 A：

$$A=\operatorname{diag}\left[I_{n-2},\begin{bmatrix}1&1\\1&1\end{bmatrix}\right],\tag{2}$$

有 $R_i(A)=0,i=1,\cdots,n-2,R_{n-1}(A)=R_n(A)=1$. 因此，当 $k\geqslant 2$ 时，$\prod\limits_{i\in\mathscr{S}}R_i(A)=0<\prod\limits_{i\in\mathscr{S}}|a_{ii}|=1$. 虽然如此，条件(1) 仍可以给出矩阵 A 秩的下界的信息. 下面介绍有关这个问题的 Shemesh 定理(见文献[4]) 的一种特殊情形，在第八章 8.1.3 小节中将研究它的推广结果.

定理 1(Shemesh) 设 $A=(a_{ij})\in\boldsymbol{M}_n(\mathbb{C}),n\geqslant 2$，且整数 k 满足 $1\leqslant k\leqslant n-1$. 若对 $\mathscr{N}$的含有 $k+1$ 个不同元素的任意子集 $\mathscr{S}$,(1)式成立，则 $\operatorname{rank}A\geqslant n-k+1$.

证明 显然地，(1)式蕴涵所有 $a_{ii}\neq 0$，且只能有 $p(\leqslant k)$个指标 i 满足 $|a_{ii}|\leqslant R_i(A)$. 当 $p=0$ 时，A 严格对角占优，因而 $\operatorname{rank}A=n\geqslant n-k+1$. 当 $p\geqslant 1$ 时，不失普遍性，假定这些 p 个指标就是前 p 个正整数，即有

$$\begin{aligned}|a_{ii}|&\leqslant R_i(A),\qquad i=1,\cdots,p,\\|a_{ii}|&>R_i(A),\qquad i=p+1,\cdots,n.\end{aligned}\tag{3}$$

考虑两种可能情形. 情形Ⅰ：$1\leqslant p<k$. 将 A 写成分块矩阵形式：

$$A=\begin{bmatrix}A_{11}&A_{12}\\A_{21}&A_{22}\end{bmatrix},\tag{4}$$

其中，$A_{22}=A[(p+1,\cdots,n)]$，其阶数为 $n-p$. 按(3)式，A_{22}严格对角占优，因而 $\operatorname{rank}A\geqslant\operatorname{rank}A_{22}=n-p\geqslant n-k+1$.

情形Ⅱ：$p=k$. 此时，(1)与(3)两式推出

$$\begin{aligned}|a_{kk}||a_{jj}|&>R_k(A)R_j(A),\qquad j=k+1,\cdots,n,\\|a_{ii}||a_{jj}|&>R_i(A)R_j(A),\qquad i,j=k,\cdots,n;i\neq j.\end{aligned}$$

现令 $\widetilde{A}_{22}=A[(k,\cdots,n)]=(\tilde{a}_{ij})$. 则上面式子推出，

$$|\tilde{a}_{ii}||\tilde{a}_{jj}|>R_i(\widetilde{A}_{22})R_j(\widetilde{A}_{22}),\qquad i,j=1,\cdots,n-k+1;i\neq j,$$

因而 6.2.1 小节的定理 1 断定，$\operatorname{rank}A\geqslant\operatorname{rank}\widetilde{A}_{22}=n-k+1$. □

我们指出，在条件(1)下定理 1 关于 $\operatorname{rank}A$ 下界的估计按照下述意义为最佳的，即总能找到满足条件(1)的矩阵 A，其秩为 $n-k+1$. 例如，由(2)式确定的 n 阶($n\geqslant 3$)矩阵 A 有秩 $n-1$. 定理 1(取 $k=2$)估计为 $\operatorname{rank}A\geqslant n-1$，又因 A 奇异，所以有 $\operatorname{rank}A=n-1$.

诚然，对 $k\geqslant 2$ 情形定理 1 不是矩阵非奇异性定理，但它也有关于特征值估计的对偶结果. 回忆一下，$\lambda\in\sigma(A)$的几何重数等于 $\dim\operatorname{Ker}(\lambda I-A)$.

推论 2 设 $A=(a_{ij})\in\boldsymbol{M}_n(\mathbb{C})(n\geqslant 2)$，$1\leqslant k\leqslant n-1$. 则 A 的每一个几何重数不

小于k的特征值λ必定位在下列$\binom{n}{k+1}$个复平面集合的并集内：

$$\left\{z \in \mathbb{C}: \prod_{i\in\mathscr{S}} |z-a_{ii}| \leqslant \prod_{i\in\mathscr{S}} R_i(A)\right\}, \tag{5}$$

其中，$\mathscr{S}\subset\mathscr{N}$的涵义如同定理 1.

该推论的证明留给读者(见习题 1).

当$A\in\boldsymbol{M}_n(\mathbb{C})$不可约时，定理 1 中条件(1)可以减弱一些，即有如下的改进结果.

定理 3 设$A=(a_{ij})\in\boldsymbol{M}_n(\mathbb{C})(n\geqslant 3)$不可约，$2\leqslant k\leqslant n-1$. 若

$$\prod_{i\in\mathscr{S}} |a_{ii}| \geqslant \prod_{i\in\mathscr{S}} R_i(A), \qquad \forall\,\mathscr{S}, \tag{6}$$

这里$\mathscr{S}\subset\mathscr{N}$涵义如同定理 1，则 $\mathrm{rank}A\geqslant n-k+1$.

证明 A不可约蕴涵$R_i(A)>0$，$\forall\, i\in\mathscr{N}$，因而由(6)式便得所有$a_{ii}\neq 0$，且只有$p(\leqslant k)$个指标$i$使得$|a_{ii}|<R_i(A)$. 当$p=0$时，$A$为不可约且对角占优的，因而$\mathrm{rank}A\geqslant n-1\geqslant n-k+1$(习题 6.1.2 题 4). 当$p\geqslant 1$时，不妨设$|a_{11}|<R_1(A)$. 现令$A_{22}=A[(2,\cdots,n)]$. 分别用$n-1$与$k-1$代替定理 1 中的$n$与$k$，则$A_{22}$满足定理 1 中的假定，因而 $\mathrm{rank}A\geqslant\mathrm{rank}A_{22}\geqslant(n-1)-(k-1)+1=n-k+1$. □

我们指出，当$n\geqslant 2$且$k=0$或$k=1$时，若不可约矩阵$A\in\boldsymbol{M}_n(\mathbb{C})$满足条件(6)，则有 $\mathrm{rank}A\geqslant n-1$. 对$k=0$情形见习题 6.1.2 题 4，对$k=1$情形，这个结论的证明将在第八章 8.1.3 小节内给出(见那里的推论 9).

现在讨论 Brualdi[3] 的矩阵非奇异性定理及其对应的排除定理，它们可以视为 6.1.1 小节定理 2，6.1.2 小节定理 8 与 6.2.1 小节定理 1 的推广. 为此，先介绍一些预备知识.

结点集合为$\mathscr{N}(n\geqslant 2)$的有向图$\Gamma$内一条**简单回路** $\boldsymbol{v}$ 是指$\mathscr{N}$的一个序列：i_1，$i_2,\cdots,i_r,i_{r+1}=i_1$，其中，$r\geqslant 2$，$i_1,\cdots,i_r$彼此不同，且$\widehat{i_1 i_2},\widehat{i_2 i_3},\cdots,\widehat{i_r i_1}$组成$\Gamma$中一条封闭的有向路. 记

$$\Gamma^{+}(i)=\{j\in\mathscr{N}: j\neq i \text{ 且有从 } i \text{ 到 } j \text{ 的有向弧线}\}, \qquad \forall\, i\in\mathscr{N}. \tag{7}$$

例如，对 6.1.2 小节图 6.3 中$\Gamma(A)$来说，$\Gamma^{+}(1)=\{2,3\}$，$\Gamma^{+}(2)=\Gamma^{+}(3)=\{1\}$；对$\Gamma(B)$来说，$\Gamma^{+}(1)=\{2\}$，$\Gamma^{+}(2)=\varnothing$.

对给定集合$\mathscr{V}$，若对它建立关系“$\prec$”使得对任意$a,b\in\mathscr{V}$，或$b\prec a$或$a\prec b$或两者皆成立，同时有$a\prec a$，$\forall\, a\in\mathscr{V}$，且$a\prec b$与$b\prec c$蕴涵$a\prec c$，其中，$a,b,c\in\mathscr{V}$，则称在$\mathscr{V}$上定义一个**前序**(preorder)“$\prec$”. 按此定义，前序仅要求自反性与传递性，不必有对称性，即$a\prec b$与$b\prec a$不必推出$a=b$. 例如，设$\mathscr{V}=\mathbb{R}$，$f:\mathbb{R}\to\mathbb{R}$为非单调的函数，定义

$$a\prec b \Leftrightarrow f(a)\leqslant f(b).$$

容易检查这是$\mathbb{R}$上一个前序，它不满足对称性，因为$f(a)=f(b)$不一定蕴涵$a=b$.

引理 4　设前序$\prec$定义在有向图Γ的结点集合$\mathcal{N}=\{1,\cdots,n\}(n\geqslant 2)$上，且$\Gamma^{\dagger}(i)\neq\varnothing$，$\forall i\in\mathcal{N}$. 则存在$\Gamma$中简单回路$\nu: i_1,i_2,\cdots,i_r,i_{r+1}=i_1$使得对每一个$j$，$1\leqslant j\leqslant r$，$i_{j+1}$为集合$\Gamma^{\dagger}(i_j)$的最大元，即$l\prec i_{j+1}$，$\forall l\in\Gamma^{\dagger}(i_j)$.

证明　从Γ的任一个结点i_1出发，取$\Gamma^{\dagger}(i_1)$的一个最大元i_2. 接着取$\Gamma^{\dagger}(i_2)$的一个最大元i_3. 若$i_3=i_1$，则简单回路$i_1,i_2,i_3=i_1$即所求；若$i_3\neq i_1$，取$\Gamma^{\dagger}(i_3)$的一个最大元i_4，照此进行下去. 记l为使得$i_l=i_m(m<l)$的最小整数，因为Γ只有有限个结点，这样的整数l一定存在. 这时$\nu: i_m,i_{m+1},\cdots,i_l=i_m$即为所求的$\Gamma$内简单回路. □

定理 5(Brualdi)　设$A=(a_{ij})\in\boldsymbol{M}_n(\mathbb{C})(n\geqslant 2)$的所有对角元素$a_{ii}\neq 0$. 若对$\Gamma(A)$的所有简单回路$\nu$，

$$\prod_{i\in\nu}|a_{ii}|>\prod_{i\in\nu}R_i(A),\tag{8}$$

则$\det A\neq 0$.

证明　假若条件(8)成立而$\det A=0$，则$A\boldsymbol{x}=\boldsymbol{0}$有解$\boldsymbol{x}=(x_1,\cdots,x_n)^{\mathrm{T}}\neq\boldsymbol{0}$. 考虑$\Gamma(A)$的有向子图$\Gamma_0(A)$，其结点集合为$\mathcal{V}=\{i\in\mathcal{N}: x_i\neq 0\}$，即$\Gamma_0(A)$是由$\Gamma(A)$去掉使$x_j=0$的那些结点$j\in\mathcal{N}$以及以这些结点为端点的$\Gamma(A)$内有向弧线得来的. 如同以前所证(见 6.2.1 小节定理 1 的证明)，$\mathcal{V}$的元素个数至少为 2. 由于$A\boldsymbol{x}=\boldsymbol{0}$且$a_{ii}\neq 0$，$\forall i\in\mathcal{N}$，故有$\Gamma_0^{\dagger}(i)\neq\varnothing$，$\forall i\in\mathcal{V}$. 在$\mathcal{V}$上定义前序$\prec$：$i\prec j\,(i,j\in\mathcal{V})$当且仅当$|x_i|\leqslant|x_j|$. 依照引理 4，$\Gamma_0(A)$有简单回路$\nu': i_1,i_2,\cdots,i_r,i_{r+1}=i_1$，使得$|x_k|\leqslant|x_{i_{j+1}}|$，$\forall k\in\Gamma_0^{\dagger}(i_j)$，$1\leqslant j\leqslant r$，从而也有$|x_k|\leqslant|x_{i_{j+1}}|$，$\forall k\in\Gamma^{\dagger}(i_j)$，$j=1,\cdots,r$. 此时，从

$$a_{i_ji_j}x_{i_j}=-\sum_{k\neq i_j}a_{i_jk}x_k,\qquad j=1,\cdots,r$$

得出，

$$\begin{aligned}0<|a_{i_ji_j}||x_{i_j}|&\leqslant\sum_{k\neq i_j}|a_{i_jk}||x_k|\\&=\sum_{k\in\Gamma^{\dagger}(i_j)}|a_{i_jk}||x_k|\leqslant\Big(\sum_{k\in\Gamma^{\dagger}(i_j)}|a_{i_jk}|\Big)|x_{i_{j+1}}|\\&=\sum_{k\neq i_j}|a_{i_jk}||x_{i_{j+1}}|\\&=R_{i_j}(A)\,|x_{i_{j+1}}|,\qquad j=1,\cdots,r.\end{aligned}\tag{9}$$

相乘之便有

$$\prod_{j=1}^{r}|a_{i_ji_j}|\prod_{j=1}^{r}|x_{i_j}|\leqslant\prod_{j=1}^{r}R_{i_j}(A)\prod_{j=1}^{r}|x_{i_{j+1}}|.$$

由于$x_{i_j}\neq 0(1\leqslant j\leqslant r)$且$i_{r+1}=i_1$，故上式蕴涵

$$\prod_{i\in\nu'}|a_{ii}|\leqslant\prod_{i\in\nu'}R_i(A),\tag{10}$$

此与(8)式矛盾.因此,$\det A\neq 0$. □

我们指出,条件 $a_{ii}\neq 0(1\leqslant i\leqslant n)$与(8)式为 6.1.1 小节(2)式与 6.2.1 小节(1)式的推广.对于 6.1.1 小节(2)式这是显然的. 6.2.1 小节(1)式蕴涵所有 $a_{ii}\neq 0$ 且至多有一个 $i_0\in\mathscr{N}$使得$|a_{i_0i_0}|\leqslant R_{i_0}(A)$.对$\Gamma(A)$的任一条简单回路 ν,若 $i_0\notin\nu$,则(8)式显然成立;若 $i_0\in\nu$,则有 $j\in\nu,j\neq i_0$,满足$|a_{i_0i_0}||a_{jj}|>R_{i_0}(A)R_j(A)$,因而(8)式也成立.因此,定理 5 确实为 Lévy-Desplanques 定理与 6.2.1 小节定理 1 的推广.

例如,对于矩阵

$$A=\begin{bmatrix}1 & 1/8 & 1/8\\ 3 & 1 & 0\\ -1 & 0 & 1\end{bmatrix},$$

它的有向图$\Gamma(A)$中只有简单回路$\{1,2,1\}$与$\{1,3,1\}$,并且$|a_{11}||a_{22}|=1>R_1(A)R_2(A)=3/4$ 与$|a_{11}||a_{33}|=1>R_1(A)R_3(A)=1/4$.因此,定理 5 断定$\det A\neq 0$.(实际上,$\det A=3/4$)但此三阶矩阵不满足 6.1.1 小节(2)式与 6.2.1 小节(1)式.

对于不可约矩阵情形,定理 5 有如下改进结果.

定理 6(Brualdi) 设 $A=(a_{ij})\in\boldsymbol{M}_n(\mathbb{C})(n\geqslant 2)$不可约.若对$\Gamma(A)$内所有简单回路 ν,

$$\prod_{i\in\nu}|a_{ii}|\geqslant\prod_{i\in\nu}R_i(A),\tag{11}$$

且至少有$\Gamma(A)$的一条简单回路使得严格不等式成立,则 $\det A\neq 0$.

证明 假若定理条件成立而 $\det A=0$,应用定理 5 的推证并注意到 A 不可约与(11)式蕴涵所有 $a_{ii}\neq 0$ 的事实,则我们有(10)式成立.顾及到(11)式,此(10)式必定为等式:

$$\prod_{i\in\nu'}|a_{ii}|=\prod_{i\in\nu'}R_i(A).\tag{12}$$

于是,(9)式中诸"$\leqslant$"号应改为等号,并且,这些 r 个等式蕴涵$|x_k|=|x_{i_{j+1}}|$,$\forall k\in\Gamma^{+}(i_j)$,$j=1,\cdots,r$.假如 ν'含有$\Gamma(A)$的所有结点,由(9)式以及刚才证明的事实可以断定,对$\Gamma(A)$的任意简单回路 ν,(11)式等式成立,此与定理假设抵触.现假定$\Gamma(A)$至少有一个结点不属于 ν'.按 6.1.2 小节定理 2,存在有向弧线从 ν'的某个结点 i_j 到$\Gamma(A)$的一个结点 $k\notin\nu'$,此表示 $k\in\Gamma^{+}(i_j)$,因而$|x_k|=|x_{i_{j+1}}|$.从引理 4 证明知道,从 i_j 与 k 出发,存在简单回路 ν'',它至少有一个结点不属于 ν',且 ν''满足该引理的要求.在定理 5 证明中,用 ν''代替 ν'可以得出,对每个 $i\in\nu''$,$|x_k|$在$\Gamma^{+}(i)$的所有结点 k 处保持不变.如果 ν'与 ν''含有$\Gamma(A)$的所有结点,应用前面的办法则可证明,对$\Gamma(A)$所有简单回路 ν,(11)式均为等式,此与定理假设抵触.否则又可以继续前面的推证,由于 A 不可约,经过有限步之后便能断定,对任意 $i\in\mathscr{N}$,$|x_k|$在$\Gamma^{+}(i)$的所有结点 k 处保持不变.于是,对所有$\Gamma(A)$的简单回路 ν,(11)式均为等

式，此与题设抵触. 因此，$\det A\neq 0$. □

与定理 5 相对应的特征值排除定理如下.

定理 7　设 $A=(a_{ij})\in \boldsymbol{M}_n(\mathbb{C})(n\geqslant 2)$ 的所有 $a_{ii}\neq 0$. 则 A 的谱 $\sigma(A)$ 包含在如下并集内：

$$\bigcup_{\nu}\left\{z\in\mathbb{C}:\prod_{i\in\nu}|z-a_{ii}|\leqslant\prod_{i\in\nu}R_i(A)\right\},\tag{13}$$

这里，ν 取遍 $\Gamma(A)$ 的所有简单回路.

与定理 6 相对应的特征值排除定理如下.

定理 8　设 $A=(a_{ij})\in \boldsymbol{M}_n(\mathbb{C})(n\geqslant 2)$ 不可约. 则 A 的任何一个特征值均不能落在并集(13)的边界上，除非它同时落在每个集合

$$\left\{z\in\mathbb{C}:\prod_{i\in\nu}|z-a_{ii}|\leqslant\prod_{i\in\nu}R_i(A)\right\}$$

的边界上，这里 ν 取遍 $\Gamma(A)$ 的所有简单回路.

容易看出，定理 7 与定理 8 分别为 6.2.1 小节定理 2 与 6.1.2 小节定理 7 的推广. 它们一般比后者提供特征值的更大排除区域.

最后我们指出，假如在定理 5 与定理 7 中将“所有 $a_{ii}\neq 0$”的假定改换为“A 的有向图 $\Gamma(A)$ **弱连接**(即 $\Gamma(A)$ 的每个结点均属于 $\Gamma(A)$ 的某个简单回路)”，则结论亦然成立. 这一点可从定理 5 的证明中直接看出(见习题 3). 有的文献将 $\Gamma(A)$ 为弱连接的矩阵 $A\in \boldsymbol{M}_n(\mathbb{C})$ 称为**弱不可约矩阵**，此与有向图强连接的矩阵为不可约矩阵的结果相呼应.

习题 6.2.2

1. 证明推论 2.

2. 证明定理 7 与定理 8.

3. 试证：若定理 5 或定理 7 中将“所有 $a_{ii}\neq 0$”的假定改换为“A 的有向图 $\Gamma(A)$ 弱连接”，则结论仍然成立.

4. 设 $A=(a_{ij})\in \boldsymbol{M}_n(\mathbb{C})$，$n\geqslant 2$，且令 $|A|\equiv(|a_{ij}|)$ 与 $M(A)\equiv(\mu_{ij})\in \boldsymbol{M}_n(\mathbb{R})$，其中，$\mu_{ij}=1$ 假如 $a_{ij}\neq 0$，且 $\mu_{ij}=0$ 假如 $a_{ij}=0$. 试证：

(1) A 不可约当且仅当 $(I+|A|)^{n-1}>O$，或等价地，$(I+M(A))^{n-1}>O$.

(2) A 弱不可约当且仅当 $B=(I+|A|)^{n-1}$ 对任意 $i\in\mathcal{N}$ 有 $b_{ij}\neq 0(j\neq i)$ 使得 $b_{ji}\neq 0$，或等价地，$\widetilde{B}=(I+M(A))^{n-1}$ 具有这个性质.

5. 设 $\Phi(\subset\mathbb{C})$ 不空. 试证：对任意 $z,w\in\Phi$，

$$z\prec w\Leftrightarrow |z|\leqslant|w|$$

定义 Φ 上一个前序. 试再举出若干前序的实例(要求它们不满足对称性).

6. 试证：$A=(a_{ij})\in \boldsymbol{M}_n(\mathbb{C})(n\geqslant 2)$ 弱不可约当且仅当不存在置换矩阵 P 使得 $P^{\mathrm{T}}AP$ 为分块上三角形矩阵且其中有一对角块为一阶矩阵.

6.3 矩阵特征值的扰动

6.3.1 特征值的连续性结果与矩阵的谱变化

为了研究矩阵的特征值连续地依赖于矩阵元素的变化的结果，需要引入两个矩阵 $A,B\in \boldsymbol{M}_n(\mathbb{C})$ 特征值集合 $\{\lambda_1,\lambda_2,\cdots\lambda_n\}$ 与 $\{\mu_1,\mu_2,\cdots,\mu_n\}$ 之间"距离"的度量. 通常应用非负实数

$$v(A,B)=\min_{\tau}\{\max_{1\leqslant j\leqslant n}|\lambda_j-\mu_{\tau(j)}|\} \tag{1}$$

表示矩阵 A 与 B 特征值集合之间的一种"距离"，称之为 **A 与 B 的特征值变差**. (1)式中 τ 取遍 $\mathcal{N}=\{1,2,\cdots,n\}$ 的所有置换. 容易验证，v 满足距离公理条件，特别有 $v(A,B)=v(B,A)$. 并且按定义，$v(A,B)\leqslant r$ 意味着存在 B 的特征值一种适当排列 $\mu_{k_1},\cdots,\mu_{k_n}$ 使得 $|\lambda_j-\mu_{k_j}|\leqslant r,j=1,\cdots,n$.

此外，也常用如下的非负实数

$$s_A(B)=\max_{1\leqslant j\leqslant n}\{\min_{1\leqslant k\leqslant n}|\lambda_k-\mu_j|\} \tag{2}$$

表示**矩阵 B 对矩阵 A 的谱变差**. 一般地，$s_A(B)\neq s_B(A)$（因而 $s_A(B)$ 不是矩阵 A 与 B 之间的一种距离）. 例如对 $A=\mathrm{diag}[-1,1]$ 与 $B=\mathrm{diag}[0,2\mathrm{i}]$，则有 $s_A(B)=\sqrt{5}$ 与 $s_B(A)=1$. 并且，$s_A(B)\leqslant r$ 意味着 $\sigma(B)\subset\bigcup_{k=1}^{n}\{z\in\mathbb{C}:|z-\lambda_k|\leqslant r\}$.

按定义，$v(A,B)\leqslant r$ 蕴涵 $s_A(B)\leqslant r$，因而我们有 $s_A(B)\leqslant v(A,B)$.

本小节基本结果如下.

定理 1(Ostrowski)　设 $A=(a_{ij}),B=(b_{ij})\in\boldsymbol{M}_n(\mathbb{C})$ 分别有特征值 $\lambda_1,\cdots,\lambda_n$ 与 $\mu_1,\cdots,\mu_n$. 令

$$m_a=\max_{i,j}(|a_{ij}|,|b_{ij}|), \tag{3}$$

$$\|B-A\|_a=\frac{1}{n}\sum_{i,j=1}^{n}|b_{ij}-a_{ij}|, \tag{4}$$

$$\delta=(n+2)m_a^{1-\frac{1}{n}}\|B-A\|_a^{1/n}. \tag{5}$$

则有

$$s_A(B)\leqslant\delta \text{ 与 } v(A,B)\leqslant(2n-1)\delta. \tag{6}$$

我们指出，由于 $\|A\|_a=\frac{1}{n}\sum_{i,j=1}^{n}|a_{ij}|$ 为 $\boldsymbol{M}_n(\mathbb{C})$ 上的广义矩阵范数（见第二章 2.2.1 小节(1) 式），故上述定理蕴涵特征值的连续性结果，即当 $B\to A$ 时，存在固定的置换 τ 使得 $\mu_{\tau(i)}\to\lambda_i,i=1,\cdots,n$. 并且更重要的是，它根据 A 与 B 元素的信息定量地给出 A 与 B 特征值之间"距离"的估计.

为证明定理 1 先考虑三个引理.

引理 2　设 $A,B,m_\alpha,\|\cdot\|_\alpha$ 与 δ 如同定理 1，$\varphi(\lambda)=\det(\lambda I-A)$ 与 $\psi(\lambda)=\det(\lambda I-B)$. 则当 $A\neq B$ 时，

$$|\varphi(\lambda)-\psi(\lambda)|<\delta^n,\ \forall\lambda\in\{z\in\mathbb{C}:|z|\leqslant nm_\alpha\}. \tag{7}$$

证明　将 A 与 B 的元素按相同次序排列为 $a_1,a_2,\cdots,a_{n^2}$ 与 $b_1,b_2,\cdots,b_{n^2}$，则对任意固定的 λ，有 n^2 个变量的多项式 p 使得

$$\varphi(\lambda)=p(a_1,a_2,\cdots,a_{n^2})\text{ 与 }\psi(\lambda)=p(b_1,b_2,\cdots,b_{n^2}).$$

令 $\Delta_k=p(a_1,\cdots,a_{k-1},a_k,b_{k+1},\cdots,b_{n^2})-p(a_1,\cdots,a_{k-1},b_k,b_{k+1},\cdots,b_{n^2}),k=1,2,\cdots,n^2$，则有

$$\varphi(\lambda)-\psi(\lambda)=\sum_{k=1}^{n^2}\Delta_k.$$

按行列式展开知道，

$$\Delta_k=\pm(a_k-b_k)\det T_k,\qquad k=1,2,\cdots,n^2,$$

式中，$T_k\in\boldsymbol{M}_{n-1}(\mathbb{C})$ 的每一行至多有一个 $\lambda-a_i$ 或 $\lambda-b_j$，其他元素为 a_m 或 b_l. 因此，只要 $|\lambda|\leqslant nm_\alpha$，$T_k$ 的每一个行向量的欧氏范数不超过 $((n-2)m_\alpha^2+(|\lambda|+m_\alpha)^2)^{1/2}<(n+2)m_\alpha$，因而按 Hadamard 不等式（见第一章 1.2.2 小节推论 12）有

$$|\det T_k|\leqslant((n+2)m_\alpha)^{n-1}.$$

这样一来，当 $|\lambda|\leqslant nm_\alpha$ 时，

$$\begin{aligned}|\varphi(\lambda)-\psi(\lambda)|&\leqslant((n+2)m_\alpha)^{n-1}\sum_{k=1}^{n^2}|a_k-b_k|\\&<(n+2)^n m_\alpha^{n-1}\|B-A\|_\alpha=\delta^n.\end{aligned}$$

□

引理 3（Rouché）　设 $\mathscr{D}$ 为 $\mathbb{C}$ 内闭域，其边界 $\partial\mathscr{D}$ 由有限多个光滑曲线所组成. 若 f 与 g 在 $\mathscr{D}$ 上单值解析，且在 $\partial\mathscr{D}$ 上满足 $|f(z)-g(z)|<|f(z)|$，则 f 与 g 在 $\mathscr{D}$ 内有相同个数的零点.

这是复分析中著名的结果，其证明可参考一般复分析教程.

引理 4　设 $A,B\in\boldsymbol{M}_n(\mathbb{C})$. 令

$$\chi_t(z)=\det(zI-((1-t)A+tB)),\qquad 0\leqslant t\leqslant 1. \tag{8}$$

若 $\mathscr{D}\subset\mathbb{C}$ 如同引理 3 所设，且 $\chi_t(z)\neq 0,\forall t\in[0,1]$ 与 $z\in\partial\mathscr{D}$，则 $\chi_t(z)$ 在 $\mathscr{D}$ 内的零点个数与 $t\in[0,1]$ 的取值无关.

证明　对给定 $t_0\in[0,1]$，令 $u(z)=\chi_{t_0}(z)$，$p=\min\limits_{z\in\partial\mathscr{D}}|\chi_{t_0}(z)|>0$ 与 $v(z)=\chi_{t_0+\varepsilon}(z)-\chi_{t_0}(z)$，其中 $t_0+\varepsilon\in[0,1]$. 由于 $\chi_t(z)$ 为闭集 $[0,1]\times\partial\mathscr{D}$ 上的连续函数，故取 $|\varepsilon|$ 足够小时可使

$$\max_{z\in\partial\mathscr{D}}|v(z)|<p.$$

这样一来，$\mathscr{D}$ 上单值解析函数 $u(z)$ 与 $v(z)$ 满足

$$|v(z)|<|u(z)|,\qquad\forall z\in\partial\mathscr{D},$$

因而按引理 3,$u(z)+v(z)=\chi_{t_0+\varepsilon}(z)$与 $u(z)=\chi_{t_0}(z)$在 $\mathscr{D}$ 内有相同个数的零点.于是在 t_0 的$|\varepsilon|$邻域内,$\chi_t(z)$在 $\mathscr{D}$ 内零点个数 $n(t)$为常数,从而 $n(t)$为$[0,1]$上的连续函数.但 $n(t)$为整数值函数,因而 $n(t)$在$[0,1]$上为常数. □

现在返回到定理 1 的证明.

定理 1 的证明 显然只需考虑 $A\neq B$ 的情形.设 $\mu_j\in\sigma(B)$,则$|\mu_j|\leqslant\|B\|_\infty\leqslant nm_\alpha$.应用引理 2,我们有

$$\prod_{i=1}^{n}|\mu_j-\lambda_i|=|\varphi(\mu_j)|=|\varphi(\mu_j)-\psi(\mu_j)|<\delta^n,$$

因而至少有一个 $\lambda_{i(j)}$满足$|\mu_j-\lambda_{i(j)}|<\delta$,这表明:$s_A(B)<\delta$,此即(6)式中第一个不等式.余下证明(6)式中第二个不等式.我们称 A 的特征值λ_i 与λ_j 相连,假如有 A 的特征值$\lambda_{\nu_1},\cdots,\lambda_{\nu_k}$使得数列 $\lambda_i,\lambda_{\nu_1},\cdots,\lambda_{\nu_k},\lambda_j$ 中相邻两数之距离都小于 2δ.将 A 的所有特征值分为若干个子集 $\mathscr{S}_1,\cdots,\mathscr{S}_s$ 使得所有相连的特征值属于同一子集,不相连的属于不同的子集.令

$$\mathscr{G}_k=\bigcup_{\lambda_i\in\mathscr{S}_k}\mathscr{D}(\lambda_i,\delta),\qquad k=1,\cdots,s,\tag{9}$$

其中,

$$\mathscr{D}(\lambda,r)=\{z\in\mathbb{C}:|z-\lambda|\leqslant r\},\lambda\in\mathbb{C}\text{与}r>0.$$

$\mathscr{G}_k$ 的边界记为$\partial\mathscr{G}_k$,它由有限多个圆弧组成.假如(8)式中函数$\chi_t(z)(0\leqslant t\leqslant1)$在并集$\bigcup\limits_{k=1}^{s}\partial\mathscr{G}_k$ 上没有零点,应用引理 4 知道,$\chi_t(z)$在每一个 $\mathscr{G}_k$ 内零点个数$n_k(t)$与 $t\in[0,1]$的取值无关,$k=1,\cdots,s$.这时,因 $\chi_0(\lambda)=\varphi(\lambda)$与 $\chi_1(\lambda)=\psi(\lambda)$,所以 $\varphi(\lambda)$与$\psi(\lambda)$在每一个 $\mathscr{G}_k$ 内有相同个数的零点.不妨假定,它们在 $\mathscr{G}_1$ 内各有 n_1 个零点,按 $\mathscr{G}_k$ 的构造知道,在 $\mathscr{G}_1$ 内,φ 的每一个零点λ_i 与ψ 的每一个零点μ_j 之间的距离满足

$$|\lambda_i-\mu_j|<(2n_1-1)\delta\leqslant(2n-1)\delta.$$

同理可证这个估计式对其他的 $\mathscr{G}_k$ 也成立.于是,有 $\mu_1,\mu_2,\cdots,\mu_n$ 的一种适当排列 $\mu_{r(1)},\mu_{r(2)},\cdots,\mu_{r(n)}$使得

$$|\lambda_i-\mu_{r(i)}|<(2n-1)\delta,i=1,\cdots,n,$$

此即 $v(A,B)<(2n-1)\delta$.因此,按前面讨论余下只要证明$\chi_t(z)(0\leqslant t\leqslant1)$在$\bigcup\limits_{k=1}^{s}\partial\mathscr{G}_k$上没有零点即可.事实上,当 $t=0$ 时,它显然成立.对某个 $t\in(0,1]$,假定 $\lambda'\in\mathbb{C}$使得$\chi_t(\lambda')=0$.现令 $\widetilde{B}=(1-t)A+tB=(\tilde{b}_{ij})$,且 $\tilde{\psi}(\lambda)=\det(\lambda I-\widetilde{B})$.则有

$$\prod_{i=1}^{n}|\lambda'-\lambda_i|=|\varphi(\lambda')|=|\varphi(\lambda')-\chi_t(\lambda')|,$$

因为$\chi_t(\lambda')=\tilde{\psi}(\lambda')=0$.按 $\widetilde{B}$ 的定义,$|\tilde{b}_{ij}|\leqslant m_\alpha$,因而 $\widetilde{B}$ 的特征值λ'满足$|\lambda'|\leqslant\|\widetilde{B}\|_\infty\leqslant nm_\alpha$.由于 $\widetilde{B}\neq A$,故根据引理 2 便得

$$\prod_{i=1}^{n}|\lambda'-\lambda_i|=|\varphi(\lambda')-\tilde{\psi}(\lambda')|<\delta^n,$$

因而必存在某个$\lambda_{i'}\in\sigma(A)$使得$|\lambda'-\lambda_{i'}|<\delta$,即$\lambda'$必属于某个$\mathscr{G}_k$内.因此,$\chi_t(z)(0<t\leqslant 1)$在$\bigcup_{k=1}^{s}\partial\mathscr{G}_k$上没有零点. □

我们指出,按上述证明,当$A\neq B$时,(6)式中二式皆为严格不等式,即$s_A(B)<\delta$与$v(A,B)<(2n-1)\delta$.

估计式(6)对一些特殊矩阵A与B来说往往偏高.但若不对A与B本身附加任何限制,则(6)式中二式右端中因子$\|B-A\|_\alpha^{1/n}$的指数$1/n$已不能进一步改善,即可以找到A与B使得$s_A(B)$或$v(A,B)$等于$\|B-A\|_\alpha^{1/n}$的常数倍.例如,对于n阶矩阵$(n\geqslant 2)$

$$A=\begin{bmatrix}0 & 1 & & & 0\\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & & 1\\ 0 & & & & 0\end{bmatrix}\ 与 B=\begin{bmatrix}0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & & 1\\ \varepsilon & & & & 0\end{bmatrix},\tag{10}$$

这里$\varepsilon>0$,有$\|B-A\|_\alpha=\varepsilon/n$,$A$的特征值均为零,$B$的特征值$\mu_1,\cdots,\mu_n$满足

$$|\mu_j-0|=\varepsilon^{1/n}=n^{1/n}\|B-A\|_\alpha^{1/n},\qquad j=1,\cdots,n.$$

从这个例子也看出,特征值的连续性结果并不意味着矩阵特征值关于矩阵元素变化的状态是稳定的或良态的,即矩阵元素的微小变化可以引起矩阵特征值相对大的变化.在前例中,矩阵A的元素a_{n1}由0变到ε,其余不变,而特征值却由0变到$\varepsilon^{1/n}$.当ε小时,$\varepsilon^{1/n}$相对于ε为一个相当大的数.譬如,$n=10$与$\varepsilon=10^{-10}$,则$\varepsilon^{1/n}=0.1$,它为ε的10^9倍(进一步讨论见文献[8]第二章).

还有一些诸如定理1给出n阶矩阵特征值随矩阵元素变化的误差界的重要结果.其中包括著名的Henrici结果.这种估计$s_A(B)$界限的方法需要应用预解式$(A-\mu I)^{-1}$的范数估计以及矩阵关于某种矩阵范数正规性偏离度的概念.在这里我们不加证明给出Henrici结果.详情参考Henrici[7]或孙继广[6]第三章第5节.

对于$A\in\boldsymbol{M}_n(\mathbb{C})$与$\boldsymbol{M}_n(\mathbb{C})$上给定的矩阵范数$\|\cdot\|_\beta$,非负实数

$$\Delta_\beta(A)\equiv\inf_M\|M\|_\beta\tag{11}$$

叫作A关于$\|\cdot\|_\beta$的**正规性偏离度**,式中M取遍A的任一酉相似变换后的上三角形矩阵形式中的严格上三角部分(见第一章1.2.1小节定理2).按此定义,正规矩阵关于$\|\cdot\|_\beta$的正规性偏离度为零,反之亦然(见习题5).

定理5(Henrici)　设$A,B\in\boldsymbol{M}_n(\mathbb{C})$,$A\neq B$,且$\|\cdot\|_\beta$为$\boldsymbol{M}_n(\mathbb{C})$上强于谱范数$\|\cdot\|_2$的任意矩阵范数,即有$\|C\|_2\leqslant\|C\|_\beta$,$\forall C\in\boldsymbol{M}_n(\mathbb{C})$.令

$$y=\Delta_\beta(A)/\|B-A\|_\beta.\tag{12}$$

则有

$$s_A(B)\leqslant\frac{y}{g(y)}\|B-A\|_\beta,\tag{13}$$

其中，$g=g(y)$为方程

$$\sum_{j=1}^{n} g^{j} = y(\geqslant 0) \tag{14}$$

的唯一非负零点. 当 $y=0$(A 为正规矩阵)时，$\lim\limits_{y\to 0^{+}} g(y)/y=1$，(13)式退化为

$$s_A(B) \leqslant \| B-A \|_{\beta}. \tag{15}$$

习题 6.3.1

1. 试应用引理 3(Rouché 结果)证明如下多项式零点关于系数变化的连续性结果：设多项式 $p(\lambda)=\lambda^n+a_{n-1}\lambda^{n-1}+\cdots+a_1\lambda+a_0$ 的系数为实数或复数，且 $p(\lambda)$零点为 $\lambda_1,\lambda_2,\cdots,\lambda_n$. 则对任意充分小 $\varepsilon>0$，存在 $\delta>0$，只要

$$\max_{0\leqslant i\leqslant n-1} | b_i - a_i | < \delta,$$

多项式 $q(\lambda)=\lambda^n+b_{n-1}\lambda^{n-1}+\cdots+b_1\lambda+b_0$ 的零点 $\mu_1,\mu_2,\cdots,\mu_n$ 便满足

$$\min_{\tau}\max_{1\leqslant j\leqslant n} | \lambda_j - \mu_{\tau(j)} | < \varepsilon,$$

这里 τ 取遍集合 $\mathcal{N}=\{1,2,\cdots,n\}$的所有置换.(注意，当取 $p(\lambda)$为 $A\in \boldsymbol{M}_n(\mathbb{C})$的特征多项式时，则得特征值连续性的结果)

2. 对于矩阵的特征向量，连续性结果一般不成立. 试以二阶矩阵

$$A_{\xi} = \begin{bmatrix} 0 & \xi \\ 0 & 0 \end{bmatrix}, \qquad \xi \in \mathbb{R}$$

为例说明之.

3. 证明：对于单重特征值对应的特征向量，连续性结果成立，即若 λ 为 $A\in \boldsymbol{M}_n(\mathbb{C})$的单重特征值，$\boldsymbol{x}\neq\boldsymbol{0}$ 为其对应的特征向量，则对 $E\in \boldsymbol{M}_n(\mathbb{C})$，$A+E$ 有特征值 $\lambda(E)$与对应的特征向量 $\boldsymbol{x}(E)$，使得当 $E\to O$ 时，$\lambda(E)\to\lambda$ 与 $\boldsymbol{x}(E)\to\boldsymbol{x}$.

4. 设 $A,B\in \boldsymbol{M}_n(\mathbb{C})$. 令 $h_A(B)=\max\limits_{0\leqslant t\leqslant 1} s_A((1-t)A+tB)$. 试证：$s_A(B)\leqslant v(A,B)\leqslant (2n-1)h_A(B)$.

5. 设 $A\in \boldsymbol{M}_n(\mathbb{C})$. 试证：若 $\Delta_F(A)$为 A 关于 $\|\cdot\|_F$ 的正规性偏离度，则

(1) A 为正规矩阵当且仅当 $\Delta_F(A)=0$.

(2) A 为正规矩阵当且仅当 $\Delta_\beta(A)=0$，这里 $\|\cdot\|_\beta$ 为任意的 $\boldsymbol{M}_n(\mathbb{C})$上矩阵范数.

(3) $\Delta_F(A)\leqslant\left(\frac{n^3-n}{12}\right)^{1/4} \| A^*A-AA^* \|_F^{1/2}$.

6.3.2 简单矩阵的特征值扰动

6.3.1 小节中的 Ostrowski 与 Henrici 结果均可视为一般矩阵 $A\in \boldsymbol{M}_n(\mathbb{C})$的特征值扰动定理，因为在那里 $B-A$ 可视为 A 的扰动，这时 $B=A+(B-A)$便是扰动后的矩阵. 但对一些特殊情形，这些定理显得不够有力. 对于简单矩阵(即可对角化矩阵)，特别是正规矩阵，应用它们特征值的许多特点，可以得到特征值扰动界限的一些特殊结果. 它们中某些结果还可推广到一般矩阵的情形，但此已超出本书

范围,有兴趣者可参阅孙继广[6]第三章.

首先,考察对角矩阵 $A=\mathrm{diag}[\lambda_1,\lambda_2,\cdots,\lambda_n]\in\boldsymbol{M}_n(\mathbb{C})$ 的特征值的扰动. 设扰动为 $E=(e_{ij})\in\boldsymbol{M}_n(\mathbb{C})$. 按 Gerschgorin 圆盘定理,扰动后的矩阵 $A+E$ 的所有特征值含在下列并集之中:

$$\bigcup_{i=1}^{n}\{z\in\mathbb{C}:|z-\lambda_i-e_{ii}|\leqslant R_i(E)\},\tag{1}$$

但此并集又含在集合

$$\bigcup_{i=1}^{n}\{z\in\mathbb{C}:|z-\lambda_i|\leqslant\sum_{j=1}^{n}|e_{ij}|\}\tag{2}$$

之中. 因此,对任意 $\hat{\lambda}\in\sigma(A+E)$,有 $\lambda_i\in\sigma(A)$ 使得

$$|\hat{\lambda}-\lambda_i|\leqslant\sum_{j=1}^{n}|e_{ij}|\leqslant\|E\|_\infty,\tag{3}$$

这表明

$$s_A(A+E)\leqslant\|E\|_\infty.\tag{4}$$

同理,用 $C_i(E)$ 代替(1)式中的 $R_i(E)$ 可证,$s_A(A+E)\leqslant\|E\|_1$.

这些简单的结果一般不能推广到非对角矩阵情形,但应用它们可以推得简单矩阵的类似结果. 事实上,若 $A\in\boldsymbol{M}_n(\mathbb{C})$ 为简单矩阵,且 $A=SDS^{-1}$,$D=\mathrm{diag}[\lambda_1,\lambda_2,\cdots,\lambda_n]$,而 $E=(e_{ij})\in\boldsymbol{M}_n(\mathbb{C})$ 为 A 的扰动,则 $\sigma(A+E)=\sigma(D+S^{-1}ES)$,因而由前一部分推证知道,对任意 $\hat{\lambda}\in\sigma(A+E)$,有 $\lambda_i\in\sigma(A)$ 使得

$$|\hat{\lambda}-\lambda_i|\leqslant\sum_{j=1}^{n}|(S^{-1}ES)_{ij}|\leqslant\|S^{-1}ES\|_\infty,\tag{5}$$

因而

$$s_A(A+E)\leqslant\|S^{-1}ES\|_\infty\leqslant\|S\|_\infty\|S^{-1}\|_\infty\|E\|_\infty=\mathrm{cond}_\infty(S)\|E\|_\infty,\tag{6}$$

式中,$\mathrm{cond}_\infty(S)=\|S\|_\infty\|S^{-1}\|_\infty$ 表示矩阵 S 关于行范数 $\|\cdot\|_\infty$ 的条件数. 同理可得,

$$s_A(A+E)\leqslant\mathrm{cond}_1(S)\|E\|_1,\tag{7}$$

式中,$\mathrm{cond}_1(S)=\|S\|_1\|S^{-1}\|_1$ 表示矩阵 S 关于列范数 $\|\cdot\|_1$ 的条件数. 不仅如此,下面定理告诉我们,上述关于 $s_A(A+E)$ 的估计对矩阵谱范数 $\|\cdot\|_2$,甚至更一般地,对 $\mathbb{C}^n$ 上绝对向量范数导出的矩阵范数 $\|\cdot\|$(见第二章 2.3.2 小节)均成立.

定理 1(Bauer-Fike)　设 $A\in\boldsymbol{M}_n(\mathbb{C})$ 为简单矩阵,且 $A=SDS^{-1}$,$D=\mathrm{diag}[\lambda_1,\lambda_2,\cdots,\lambda_n]$,而 $E\in\boldsymbol{M}_n(\mathbb{C})$ 为 A 的扰动. 若 $\|\cdot\|$ 为 $\mathbb{C}^n$ 上绝对向量范数导出的矩阵范数,则对任意 $\hat{\lambda}\in\sigma(A+E)$,有 $\lambda_i\in\sigma(A)$ 使得

$$|\hat{\lambda}-\lambda_i|\leqslant\mathrm{cond}(S)\|E\|,\tag{8}$$

这里,$\mathrm{cond}(S)=\|S\|\|S^{-1}\|$ 表示矩阵 S 关于 $\|\cdot\|$ 的条件数,并且,

$$s_A(A+E) \leqslant \operatorname{cond}(S)\|E\|. \tag{9}$$

证明 令 $B=D+S^{-1}ES$,则显然 $\sigma(A+E)=\sigma(B)$. 对任意 $\hat{\lambda}\in\sigma(B)$,若 $D-\hat{\lambda}I$ 奇异,则 $\hat{\lambda}$ 为 A 的某个特征值 λ_i,因而 $|\hat{\lambda}-\lambda_i|=0$. 若 $D-\hat{\lambda}I$ 非奇异,则

$$B-\hat{\lambda}I=(D-\hat{\lambda}I)(I+(D-\hat{\lambda}I)^{-1}S^{-1}ES),$$

因而 $I+(D-\hat{\lambda}I)^{-1}S^{-1}ES$ 奇异,于是,$-1\in\sigma((D-\hat{\lambda}I)^{-1}S^{-1}ES)$. 再根据第二章 2.2.2 小节定理 4 与 2.3.2 小节定理 4,

$$\begin{aligned}1 &\leqslant \|(D-\hat{\lambda}I)^{-1}S^{-1}ES\| \leqslant \|(D-\hat{\lambda}I)^{-1}\|\,\|S^{-1}ES\| \\ &= (\min_i|\lambda_i-\hat{\lambda}|)^{-1}\|S^{-1}ES\|.\end{aligned}$$

因此,

$$\min_i|\lambda_i-\hat{\lambda}| \leqslant \|S^{-1}ES\| \leqslant \operatorname{cond}(S)\|E\|,\ \forall\hat{\lambda}\in\sigma(B),$$

由此即得(8)式与(9)式. □

应用 6.1.1 小节中 Gerschgorin 圆盘定理的推证方法,可得到定理 1 的如下进一步结果(证明留给读者,见习题 4).

定理 2 设 A,S,E 与 $\|\cdot\|$ 同定理 1. 若 λ_i 为 A 的 m 重特征值,且圆盘 $\mathscr{D}=\{z\in\mathbb{C}:|z-\lambda_i|\leqslant\|S^{-1}ES\|\}$与圆盘$\{z\in\mathbb{C}:|z-\lambda_k|\leqslant\|S^{-1}ES\|\}$不相交,$\forall\lambda_k\neq\lambda_i$,则 $\mathscr{D}$ 恰含有 $A+E$ 的 m 个特征值.

假如已经求得 $A\in\boldsymbol{M}_n(\mathbb{C})$的近似特征值 λ 与对应的近似特征向量 $\boldsymbol{x}(\neq\boldsymbol{0})$,为了检验它们的准确性,自然希望考虑残差向量 $\boldsymbol{r}=A\boldsymbol{x}-\lambda\boldsymbol{x}$. 若 $\boldsymbol{r}=\boldsymbol{0}$,则 λ 与 $\boldsymbol{x}$ 显然是精确的,但若 $\boldsymbol{r}\neq\boldsymbol{0}$,即使 $\|\boldsymbol{r}\|$ 只是小的量,λ 与 A 的特征值可能差异甚大. 例如,二阶矩阵

$$A=\begin{bmatrix}2 & -10^{10}\\ 0 & 2\end{bmatrix}$$

有特征值 2,2. 取 $\boldsymbol{x}=(1,10^{-10})^{\mathrm{T}}$ 与 $\lambda=1$,则 $\|\boldsymbol{r}\|_\infty=\|A\boldsymbol{x}-\lambda\boldsymbol{x}\|_\infty=10^{-10}$,而 $|\lambda-2|=1$. 但是,对于简单矩阵,下面结果告诉我们,可以用残差向量的大小来估计近似特征值的误差.

定理 3 设 $A\in\boldsymbol{M}_n(\mathbb{C})$为简单矩阵,$A=SDS^{-1}$,$D=\operatorname{diag}[\lambda_1,\lambda_2,\cdots,\lambda_n]$,且 $\|\cdot\|$ 为任意绝对向量范数或它导出的矩阵范数. 若 $\|\boldsymbol{r}\|=\|A\boldsymbol{x}-\lambda\boldsymbol{x}\|\leqslant\varepsilon$,且 $\|\boldsymbol{x}\|=1$,则有

$$\min_{1\leqslant i\leqslant n}|\lambda_i-\lambda|\leqslant\varepsilon\operatorname{cond}(S), \tag{10}$$

式中,$\operatorname{cond}(S)$涵义同定理 1.

证明 不妨设 $\lambda\notin\sigma(A)$,否则(10)式自然成立. 这时

$$\boldsymbol{r}=A\boldsymbol{x}-\lambda\boldsymbol{x}=S(D-\lambda I)S^{-1}\boldsymbol{x},$$

因而应用第二章 2.3.2 小节定理 4 得出

$$1=\|\boldsymbol{x}\|=\|S(D-\lambda I)^{-1}S^{-1}\boldsymbol{r}\|\leqslant\varepsilon\|S\|\,\|S^{-1}\|/\min_i|\lambda_i-\lambda|,$$

此相当于(10)式. □

定理 1 与定理 2 都说明,在简单矩阵 A 的特征值扰动分析中,$S^{-1}ES$ 的某种矩阵范数为关键的因素,这里非奇异矩阵 S 使得 $S^{-1}AS=D=\mathrm{diag}[\lambda_1,\lambda_2,\cdots,\lambda_n]$. 记 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n$ 为 S 的 n 个列向量,$s_1^{-1}\boldsymbol{y}_1^{\mathrm{T}},\cdots,s_n^{-1}\boldsymbol{y}_n^{\mathrm{T}}$ 为 S^{-1} 的 n 个行向量,这里 $s_1,\cdots,s_n$是合适的非零复数使得 $\|\boldsymbol{x}_i\|_2=\|\boldsymbol{y}_i\|_2,i=1,2,\cdots,n$. 则 $\boldsymbol{x}_i$ 与 $\boldsymbol{y}_i$ 分别为矩阵$A\in\boldsymbol{M}_n(\mathbb{C})$的右与左特征向量,$i=1,2,\cdots,n$. 由于 $S^{-1}S=I$,故有

$$\boldsymbol{y}_i^{\mathrm{T}}\boldsymbol{x}_j=s_j\delta_{ij}. \tag{11}$$

现假定 λ_i 为 A 的单重特征值. 按定理 2,对充分小 $\|E\|_\infty$,$\mathscr{D}=\{z\in\mathbb{C}:|z-\lambda_i|\leqslant\|S^{-1}ES\|_\infty\}$恰含有 $A+E$ 的一个特征值μ_j,因而

$$|\lambda_i-\mu_j|\leqslant\sum_{k=1}^n|(S^{-1}ES)_{ik}|=\sum_{k=1}^n|\boldsymbol{y}_i^{\mathrm{T}}E\boldsymbol{x}_k|/|s_i|. \tag{12}$$

(注意,在定理 2 证明中实际上可得到 $\widetilde{\mathscr{D}}=\{z\in\mathbb{C}:|z-\lambda_i|\leqslant\sum\limits_{k=1}^n|(S^{-1}ES)_{ik}|\}$恰含有 $A+E$ 的 m 个特征值的结论)在这个估计中,数量$|s_i|=|\boldsymbol{y}_i^{\mathrm{T}}\boldsymbol{x}_i|$起着基本的作用:它越大,特征值 λ_i 的扰动一般越小. 因此,对简单矩阵 $A\in\boldsymbol{M}_n(\mathbb{C})$,可用 $\|\boldsymbol{x}_i\|_2^2|\boldsymbol{y}_i^{\mathrm{T}}\boldsymbol{x}_i|^{-1}$定义它的特征值 λ_i 的条件数. 受此启发,我们对一般方阵的单重特征值引入以下定义.

定义 4　设 λ 为 $A\in\boldsymbol{M}_n(\mathbb{C})$的**单重**特征值,$\boldsymbol{x}$ 与 $\boldsymbol{y}$ 分别为对应的右与左特征向量,且 $\|\boldsymbol{x}\|_2=\|\boldsymbol{y}\|_2=1$. 则称$|\boldsymbol{y}^{\mathrm{T}}\boldsymbol{x}|^{-1}$为**特征值 λ 的条件数**.

我们指出,若去掉 $\|\boldsymbol{x}\|_2=\|\boldsymbol{y}\|_2=1$ 的限制,则定义 4 中的单重特征值 λ 的条件数可由

$$s=\frac{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{x}}{\|\boldsymbol{x}\|_2\|\boldsymbol{y}\|_2}$$

的绝对值倒数 $1/|s|$来确定. 有的文献中选 $\|\boldsymbol{x}\|_2=|\boldsymbol{y}^{\mathrm{T}}\boldsymbol{x}|=1$,因而定义 $\|\boldsymbol{y}\|_2$ 为单重特征值 λ 的条件数. 根据 Cauchy 不等式,此特征值条件数总不小于 1.

为了说明定义 4 中$|\boldsymbol{y}^{\mathrm{T}}\boldsymbol{x}|^{-1}$确实与特征值 λ 的状态有关,我们考察当 $E\to O$ 时定理 2 的极限情形. 设 $A\in\boldsymbol{M}_n(\mathbb{C})$为简单矩阵,且 $S^{-1}AS=D=\mathrm{diag}[\lambda_1,\cdots,\lambda_n]$. 若 λ_i 为 A 的单重特征值,其对应的右与左特征向量分别为 $\boldsymbol{x}$ 与 $\boldsymbol{y}$,$\|\boldsymbol{x}\|_2=\|\boldsymbol{y}\|_2=1$,且 $E=\varepsilon C$ 为扰动,其中 $C\in\boldsymbol{M}_n(\mathbb{C})$给定,取 ε 充分小使得圆盘 $\mathscr{D}=\{z\in\mathbb{C}:|z-\lambda_i|\leqslant\|S^{-1}ES\|_\infty\}$恰含 $A+E$ 的一个特征值$\mu(\varepsilon)$,则有

$$\mu'(0)=\lim_{\varepsilon\to0}\frac{\mu(\varepsilon)-\lambda_i}{\varepsilon}=\frac{\boldsymbol{y}^{\mathrm{T}}C\boldsymbol{x}}{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{x}}. \tag{13}$$

事实上,令 $\boldsymbol{x}(\varepsilon)$为对应于$\mu(\varepsilon)$的 $A+E$ 的特征向量. 于是

$$A(\boldsymbol{x}(\varepsilon)-\boldsymbol{x})+\varepsilon C\boldsymbol{x}(\varepsilon)=(\mu(\varepsilon)-\lambda_i)\boldsymbol{x}(\varepsilon)+\lambda_i(\boldsymbol{x}(\varepsilon)-\boldsymbol{x}),$$

两边左乘以 $\boldsymbol{y}^{\mathrm{T}}$ 便得

$$\varepsilon\boldsymbol{y}^{\mathrm{T}}C\boldsymbol{x}(\varepsilon)=(\mu(\varepsilon)-\lambda_i)\boldsymbol{y}^{\mathrm{T}}\boldsymbol{x}(\varepsilon). \tag{14}$$

根据习题 6.3.1 题 3,可选择适当的 $x(\varepsilon)$ 使得 $\lim\limits_{\varepsilon\to 0}x(\varepsilon)=x$. 让 $\varepsilon\to 0$,由(14)式即得(13)式.

现在考虑 A 与扰动 E 均为 Hermite 矩阵的情形. 基本问题还是求当将 A 代以 $A+E$ 时,特征值扰动的界.

定理 5(Weyl) 设 $A,E\in \boldsymbol{M}_n(\mathbb{C})$ 均为 Hermite 矩阵,且 $A,A+E$ 与 E 的特征值按不增次序排列分别为 $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_n,\mu_1\geqslant\mu_2\geqslant\cdots\geqslant\mu_n$,与 $\nu_1\geqslant\nu_2\geqslant\cdots\geqslant\nu_n$. 则有

$$\nu_n\leqslant\mu_k-\lambda_k\leqslant\nu_1,k=1,2,\cdots,n, \tag{15}$$

证明 按第一章 1.2.1 小节定理 6 后面说明,E 的 Rayleigh 商有如下估计:

$$\nu_n\leqslant\frac{\boldsymbol{x}^*E\boldsymbol{x}}{\boldsymbol{x}^*\boldsymbol{x}}\leqslant\nu_1,\qquad \forall\,\boldsymbol{0}\neq\boldsymbol{x}\in\mathbb{C}^n.$$

并且,若 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n$ 表示对应于 $\lambda_1,\cdots,\lambda_n$ 的 A 的标准正交特征向量,$\mathscr{R}$ 表示由 $\boldsymbol{x}_k,\cdots,\boldsymbol{x}_n$ 生成的$\mathbb{C}^n$ 子空间,则有

$$\mu_k\leqslant\max_{\substack{\boldsymbol{u}\in\mathscr{R}\\ \boldsymbol{u}^*\boldsymbol{u}=1}}\boldsymbol{u}^*(A+E)\boldsymbol{u}\leqslant\max_{\substack{\boldsymbol{u}\in\mathscr{R}\\ \boldsymbol{u}^*\boldsymbol{u}=1}}\boldsymbol{u}^*A\boldsymbol{u}+\max_{\substack{\boldsymbol{u}\in\mathscr{R}\\ \boldsymbol{u}^*\boldsymbol{u}=1}}\boldsymbol{u}^*E\boldsymbol{u}\leqslant\lambda_k+\nu_1,$$
$$k=1,\cdots,n, \tag{16}$$

这里,我们引用了第一章 1.2.1 小节中 Courant-Fischer 定理. 另一方面,若令 $\widetilde{E}=-E$,则其特征值为 $-\nu_n\geqslant-\nu_{n-1}\geqslant\cdots\geqslant-\nu_1$. 由于 $A=(A+E)+\widetilde{E}$,故按(16)式(其中 A 与 E 分别换为 $A+E$ 与 $\widetilde{E}$)有

$$\lambda_k\leqslant\mu_k-\nu_n,\qquad k=1,\cdots,n.$$ □

从(15)式直接推出,对 $k=1,\cdots,n$,

$$|\mu_k-\lambda_k|\leqslant\max\{\nu_1,|\nu_n|\}=\|E\|_2.$$

因此有如下结论,其中(18)式显然是(17)式的直接推论.

定理 6(Weyl-Lidskii) 设 $A,E\in\boldsymbol{M}_n(\mathbb{C})$ 为 Hermite 矩阵,且 A 与 $A+E$ 的特征值分别为 $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_n$ 与 $\mu_1\geqslant\mu_2\geqslant\cdots\geqslant\mu_n$. 则有

$$|\mu_k-\lambda_k|\leqslant\|E\|_2,\qquad k=1,\cdots,n, \tag{17}$$

$$v(A,A+E)\leqslant\|E\|_2. \tag{18}$$

我们指出,定理 6 表明 Hermite 矩阵 A 的小的 Hermite 扰动 E 只引起 A 的特征值相应的小变化. 在这种意义之下,Hermite 矩阵的特征值是**稳定的**(若按定义 4 前面说明,当 $A=A^*$ 时,A 的左特征向量 $\boldsymbol{y}$ 恰为对应的右特征向量 $\boldsymbol{x}$ 的共轭:$\boldsymbol{y}=\bar{\boldsymbol{x}}$. 假如 $\|\boldsymbol{x}\|_2=\|\boldsymbol{y}\|_2=1$,则 $|\boldsymbol{y}^{\mathrm{T}}\boldsymbol{x}|^{-1}=|\boldsymbol{x}^*\boldsymbol{x}|^{-1}=\|\boldsymbol{x}\|_2^{-1}=1$. 因此,$A$ 的单重特征值的条件数均为 1).

定理 6 中(18)式蕴涵 $s_A(A+E)\leqslant\|E\|_2$,此估计也可以从定理 1 中(9)式推出,因为 A 这时酉相似于一个对角矩阵,而对任意酉矩阵 U,$\mathrm{cond}_2(U)=1$. 注意,结果(17)式显然比(18)式蕴涵更强的内容.

由于 $\|E\|_F^2=\mathrm{tr}(E^2)=\nu_1^2+\cdots+\nu_n^2\leqslant n\|E\|_2^2$,故下面结果可视为定理 6 的补

充.(其证明见后面推论 11)

定理 7(Hoffman-Wielandt)　设 $A,A+E$ 与 E 皆为 n 阶 Hermite 矩阵,它们特征值排序如定理 5. 则有

$$\left(\sum_{k=1}^{n}(\mu_k-\lambda_k)^2\right)^{1/2}\leqslant\|E\|_F. \tag{19}$$

Hermite 矩阵特征值扰动的另一个有关结果,涉及向量集合的凸包. 回忆一下,一个向量集合的凸包意指含该集合的最小凸集.

定理 8(Lidskii)　设 $A,A+E$ 与 E 皆为 n 阶 Hermite 矩阵,它们特征值排序如定理 5. 若令 $\boldsymbol{\lambda}=(\lambda_1,\cdots,\lambda_n)^{\mathrm{T}},\mu=(\mu_1,\cdots,\mu_n)^{\mathrm{T}}$ 与 $\boldsymbol{v}=(\nu_1,\cdots,\nu_n)^{\mathrm{T}}$,则 n 维实向量 $\boldsymbol{\mu}$ 位于 $\boldsymbol{\lambda}+P\boldsymbol{v}$ 组成的向量集合的凸包中,这里,P 取遍所有 n 阶置换矩阵.

我们将略去上述定理的证明. 在此只考虑它的一个应用例子. 假定在定理 6 中,$n=2,\lambda_1=2,\lambda_2=1,E$ 的特征值为:$\nu_1=2\varepsilon,\nu_2=-\varepsilon,\varepsilon>0$. 于是,$\|E\|_2=|\nu_1|=2\varepsilon$. 按公式(17),则有

$$|\mu_1-2|\leqslant2\varepsilon\quad 与\quad |\mu_2-1|\leqslant2\varepsilon,$$

这表明向量 $\boldsymbol{\mu}=(\mu_1,\mu_2)^{\mathrm{T}}$ 一定都落在中心在(2,1),边长为 4ε 的正方形内(其边与坐标轴平行). 若应用定理 8,则 $\boldsymbol{\mu}$ 落在$\{\boldsymbol{\lambda}+P\boldsymbol{v}\}$的凸包内,它是过两点$(2+2\varepsilon,1-\varepsilon)$与$(2-\varepsilon,1+2\varepsilon)$之间的线段. 因此,对于这个实例,定理 8 比起定理 6 给出向量 $\boldsymbol{\mu}$ 更为精确的位置估计.

最后,考虑 $A\in\boldsymbol{M}_n(\mathbb{C})$为正规矩阵的情形. 主要讨论当扰动后矩阵 $B\in\boldsymbol{M}_n(\mathbb{C})$ 也为正规矩阵时,A 与 B 特征值偏差的估计.

下文需要应用双随机矩阵的 Birkhoff 定理. 我们将在第七章 7.3.2 小节定理 3 给出它的证明.

定理 9(Birkhoff)　设 $S=(s_{ij})\in\boldsymbol{M}_n(\mathbb{R})$为双随机矩阵(即所有 $s_{ij}\geqslant0$ 且所有行和与列和都等于 1). 则 S 可以表示为 n 阶置换矩阵 $P_i(i=1,2,\cdots,n!)$的凸组合:

$$S=\sum_{i=1}^{n!}\sigma_iP_i,\qquad\sum_{i=1}^{n!}\sigma_i=1,\qquad\sigma_i\geqslant0,i=1,2,\cdots,n!. \tag{20}$$

定理 10(Hoffman-Wielandt)　设 $A,B\in\boldsymbol{M}_n(\mathbb{C})$均为正规矩阵,$\{\lambda_1,\cdots,\lambda_n\}$与$\{\mu_1,\cdots,\mu_n\}$分别为 A 与 B 按某种相同的次序排列的特征值. 则存在 $\mathcal{N}=\{1,\cdots,n\}$ 的一种置换 π 使得

$$\left(\sum_{i=1}^{n}|\mu_{\pi(i)}-\lambda_i|^2\right)^{1/2}\leqslant\|B-A\|_F. \tag{21}$$

证明　设 $D=\mathrm{diag}[\lambda_1,\cdots,\lambda_n]$与 $\hat{D}=\mathrm{diag}[\mu_1,\cdots,\mu_n]$,且 $U,V\in\mathscr{U}_n$ 使得 $A=UDU^*$ 与 $B=V\hat{D}V^*$. 这时由于 $\|\cdot\|_F$ 为酉不变的,令 $W=U^*V=(w_{ij})$我们有

$$\begin{aligned}\|B-A\|_F^2&=\|V\hat{D}V^*-UDU^*\|_F^2=\|W\hat{D}W^*-D\|_F^2\\&=\mathrm{tr}((W\hat{D}W^*-D)(W\hat{D}W^*-D)^*)\end{aligned}$$

$$= \sum_{i=1}^{n}(|\mu_i|^2+|\lambda_i|^2)-2\mathrm{Retr}(W\hat{D}W^*D^*)$$

$$= \sum_{i=1}^{n}(|\mu_i|^2+|\lambda_i|^2)-g(W), \tag{22}$$

式中，

$$g(W)=2\mathrm{Retr}(W\hat{D}W^*D^*)=2\sum_{i,j=1}^{n}\mathrm{Re}(\lambda_i\bar{\mu}_j)\,|w_{ij}|^2. \tag{23}$$

由(22)式看出

$$\|B-A\|_F^2 \geqslant \sum_{i=1}^{n}(|\mu_i|^2+|\lambda_i|^2)-\max_{W\in\mathcal{U}_n} g(W). \tag{24}$$

我们将证明：上式右端恰为(21)式的左端值之平方，其中 π 为 $\mathcal{N}$ 的某个置换.

记 $S=(|w_{ij}|^2)\in \boldsymbol{M}_n(\mathbb{R})$. 则 S 为双随机矩阵. 从(23)式推得

$$\max_{W\in\mathcal{U}_n} g(W)\leqslant 2\max\left\{\sum_{i,j=1}^{n}\mathrm{Re}(\lambda_i\bar{\mu}_j)s_{ij}\ :\ S=(s_{ij})\in\Omega_n\right\},$$

式中，Ω_n 表示 n 阶双随机矩阵的集合. 但出现在上式右端花括号内的函数为 Ω_n 上的线性函数，记为

$$f(S)=\sum_{i,j=1}^{n}\mathrm{Re}(\lambda_i\bar{\mu}_j)s_{ij}, \qquad \forall S\in\Omega_n.$$

按定理 9，我们有 $\sigma_k\geqslant 0$，$\sum_{k=1}^{n!}\sigma_k=1$ 使得

$$f(S)=f\left(\sum_{k=1}^{n!}\sigma_k P_k\right)=\sum_{k=1}^{n!}\sigma_k f(P_k), \qquad \forall S\in\Omega_n.$$

设 $P=(\delta_{\pi(i),j})\in\Omega_n$（其中 π 为 $\mathcal{N}$ 的某置换）使得 $f(P)=\max_{1\leqslant k\leqslant n!} f(P_k)=\max_{S\in\Omega_n} f(S)$.
注意到 $P\in\mathcal{U}_n$，便得

$$\max_{w\in\mathcal{U}_n} g(W)=g(P)=2\sum_{i=1}^{n}\mathrm{Re}(\lambda_i\bar{\mu}_{\pi(i)})=2f(P).$$

代入(24)式，最后推出，

$$\|B-A\|_F^2\geqslant\sum_{i=1}^{n}(|\lambda_i|^2+|\mu_i|^2-2\mathrm{Re}(\lambda_i\bar{\mu}_{\pi(i)}))=\sum_{i=1}^{n}|\mu_{\pi(i)}-\lambda_i|^2. \quad \square$$

我们指出，当 A 与 B 不同时为正规矩阵时定理 10 结论不真. 例如，对于

$$A=\begin{bmatrix}0&0\\0&4\end{bmatrix} \quad 与 \quad B=\begin{bmatrix}-1&-1\\1&1\end{bmatrix},$$

$\|A-B\|_F^2=12$，但对任意置换 π，$\sum|\mu_{\pi(i)}-\lambda_i|^2=16$.

当 $A\in\boldsymbol{M}_n(\mathbb{C})$ 为 Hermite 矩阵，而 $B\in\boldsymbol{M}_n(\mathbb{C})$ 为正规矩阵时，定理 10 的结果可以得到如下的改进.

推论 11　设 $A\in \boldsymbol{M}_n(\mathbb{C})$ 为 Hermite 矩阵，$B\in \boldsymbol{M}_n(\mathbb{C})$ 为正规矩阵，它们特征值 $\{\lambda_1,\cdots,\lambda_n\}$ 与 $\{\mu_1,\cdots,\mu_n\}$ 分别按 $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_n$ 与 $\mathrm{Re}\mu_1\geqslant\mathrm{Re}\mu_2\geqslant\cdots\geqslant\mathrm{Re}\mu_n$ 的次序排列. 则

$$\Big(\sum_{i=1}^{n}|\mu_i-\lambda_i|^2\Big)^{1/2}\leqslant \|B-A\|_F. \tag{25}$$

证明　按定理 10，有 $\mathcal{N}$ 的一种置换 π 使得

$$\Big(\sum_{i=1}^{n}|\mu_{\pi(i)}-\lambda_i|^2\Big)^{1/2}\leqslant \|B-A\|_F. \tag{26}$$

假如 π 为 $\mathcal{N}$ 的恒等置换，即 $\pi(i)=i,i=1,\cdots,n$，则上式即为(25)式. 假如 π 不是 $\mathcal{N}$ 的恒等置换(即有某 $i\in\mathcal{N}$ 使 $\pi(i)\neq i$)，则存在 $k\in\mathcal{N}$ 使得 $\mathrm{Re}\mu_{\pi(k)}<\mathrm{Re}\mu_{\pi(k+1)}$. 但因 $\lambda_k\geqslant\lambda_{k+1}$，并且，

$$\begin{aligned}|\mu_{\pi(k)}-\lambda_k|^2+|\mu_{\pi(k+1)}-\lambda_{k+1}|^2=&|\mu_{\pi(k+1)}-\lambda_k|^2+|\mu_{\pi(k)}-\lambda_{k+1}|^2\\&+2(\lambda_k-\lambda_{k+1})(\mathrm{Re}\mu_{\pi(k+1)}-\mathrm{Re}\mu_{\pi(k)}),\end{aligned}$$

所以有

$$|\mu_{\pi(k)}-\lambda_k|^2+|\mu_{\pi(k+1)}-\lambda_{k+1}|^2\geqslant|\mu_{\pi(k+1)}-\lambda_k|^2+|\mu_{\pi(k)}-\lambda_{k+1}|^2.$$

因此，交换 B 的特征值 $\mu_{\pi(k)}$ 与 $\mu_{\pi(k+1)}$ 的位置不增大(26)式左端的值. 按此进行有限次可以使 $\mu_{\pi(1)},\cdots,\mu_{\pi(n)}$ 依实部递降的顺序排列为 $\mu_1,\cdots,\mu_n$，由于每一次变换都不增大(26)式左端的值，因而(25)式成立. □

特别地，若 $B\in \boldsymbol{M}_n(\mathbb{C})$ 也为 Hermite 矩阵，则推论 11 便退化为前面的定理 7.

习题 6.3.2

1. 设 $A\in \boldsymbol{M}_n(\mathbb{C})$ 为简单矩阵，且 $\|\cdot\|$ 为 $\mathbb{C}^n$ 上某绝对向量范数导出的矩阵范数. 记 $\nu(A)=\inf\limits_{S}(\|S\|\,\|S^{-1}\|)$，这里，$S$ 非奇异使得 $S^{-1}AS$ 为对角矩阵. 试证：

$$\|A\|/\nu(A)\leqslant\max_{\lambda\in\sigma(A)}|\lambda|\leqslant\|A\|,$$
$$\mathrm{glb}(A)\leqslant\min_{\lambda\in\sigma(A)}|\lambda|\leqslant\nu(A)\mathrm{glb}(A).$$

2. 设 $A,B\in \boldsymbol{M}_n(\mathbb{C})$，$A$ 为简单矩阵有特征值 $\lambda_1,\cdots,\lambda_n$，且 $\|\cdot\|$ 为 $\mathbb{C}^n$ 上某绝对向量范数导出的矩阵范数. 试证：$A+B$ 的任何特征值 μ 必定属于下列 n 个圆盘的并集：

$$\{z\in\mathbb{C}:|z-\lambda_i|\leqslant\|B\|\nu(A)\},\qquad i=1,\cdots,n,$$

这里，$\nu(A)$ 涵义同题 1.

3. 设 $A,B,\|\cdot\|$ 如题 2，但 B 也为简单矩阵有特征值 $\mu_1,\cdots,\mu_n$. 试证：$A+B$ 的任何特征值 μ 必定属于下列 n 个圆盘的并集：

$$\{z\in\mathbb{C}:|z-\lambda_i|\leqslant\nu(A,B)\rho(B)\},\qquad i=1,\cdots,n,$$

这里，$\nu(A,B)=\inf\limits_{S,T}(\|S^{-1}T\|\,\|T^{-1}S\|)$，式中 S 与 T 非奇异，分别使得 $S^{-1}AS$ 与 $T^{-1}BT$ 为对角矩阵.

4. 证明定理 2.

5. 设 $H\in \boldsymbol{M}_n(\mathbb{C})$ 为 Hermite 矩阵，有特征值 $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_n$. 试证：对 $k=1,\cdots,n$，

$$\max_{UU^*=I_k} \operatorname{tr}(UHU^*) = \sum_{j=1}^{k} \lambda_j,$$

$$\min_{UU^*=I_k} \operatorname{tr}(UHU^*) = \sum_{j=1}^{k} \lambda_{n-j+1},$$

这里，U 取遍满足 $UU^* = I_k$ 的 $k \times n$ 复矩阵.

6. 设 $A, A+E$ 与 E 皆为 n 阶 Hermite 矩阵，它们的特征值排序如定理 5. 试证：

$$\sum_{j=1}^{k} \lambda_j + \sum_{j=1}^{k} \nu_{n-j+1} \leqslant \sum_{j=1}^{k} \mu_j \leqslant \sum_{j=1}^{k} \lambda_j + \sum_{j=1}^{k} \nu_j, \qquad k = 1, \cdots, n.$$

7. 试证：在推论 11 假定条件下，如令 $E=B-A$，则有推论 11 的如下改进结果：

$$\left(\sum_{i=1}^{n} (\operatorname{Re}\mu_i - \lambda_i)^2\right)^{1/2} \leqslant \| \operatorname{Re}E \|_F, \left(\sum_{i=1}^{n} (\operatorname{Im}\mu_i)^2\right)^{1/2} \leqslant \| \operatorname{Im}E \|_F.$$

参考文献

[1] Householder A S. The Theory of Matrices in Numerical Analysis. Blaisdell, Boston, 1964(前三章有中译本，孙家昶等译. 数值分析中的矩阵论. 北京：科学出版社，1986), New York, Reprinted by Dover, 1975

[2] Marcus M, Minc H. A Survey of Matrix Theory and Matrix Inequalities. Allyn and Bacon, Rockleigh, New York: New Jersey, 1964, Reprinted by Dover, 1992

[3] Brualdi R. Matrices, eigenvalues, and directed graphs. Linear and Multilinear Algebra, 1982 (11), 143～165

[4] Dan Shemesh. On a theorem of A. Ostrowski. Linear and Multilinear Algebra, 1977(5), 139～143

[5] Gong-ning Chen(陈公宁). Note on lower bounds for the rank of a matrix. Linear Algebra Appl. 1983(55), 125～132

[6] 孙继广. 矩阵扰动分析. 北京：科学出版社，1987

[7] Henrici P. Bounds for iterates, inverses, spectral variation and fields of values of non-normal matrices. Numer. Math., 1962(4), 24～39

[8] Kato T. A Short Introduction to Perturbation Theory for Linear Operators. New York: Springer-Verlag, 1982

[9] Baumgartel H. Analytic Perturbation Theory for Matrices and Operators. Basel: Birkhauser, 1985

[10] Gohberg I., Lancaster P., Rodman L. Invariant Subspaces of Matrices with Applications. New York: John Wiley and Sons, 1986

[11] Rellich F. Perturbation Theory of Eigenvalue Problems. New York: Gordon and Breach, 1969

[12] Wilkinson J H. The Algebraic Eigenvalue Problem. Oxford: Oxford Univ. Press, 1965

[13] Stewart G W, Sun J. Matrix Perturbation Theory. San Diego: Academic Press, 1990

[14] Parlett B N. The Symmetric Eigenvalue Problem. New Jersey: Prentice-Hall, Inc, Englewood Cliffs, 1997

第七章 非负矩阵理论

在 5.2.3 小节中我们已经引入非负矩阵与正矩阵的概念与记号. 一个矩阵 A 叫作非负(正)的,假如它的所有元素为非负(正)的实数,记为 $A\geqslant O(A>O)$. 非负方阵有一些值得注意的谱性质. 对于正方阵,Perron 在 1907 年首先发现这些谱性质,1908～1912 年 Frobenius 扩大并推广 Perron 结果到非负矩阵特别到非负不可约矩阵情形. 在 1973～1975 年,对可约矩阵情形的研究也取得令人满意的成果. 7.1 节主要介绍非负不可约矩阵(包括正方阵)的 Perron-Frobenius 理论,7.2 节讨论一般非负方阵的 Perron-Frobenius 理论的古典结果与它的推广,最后在 7.3 节中我们介绍两种特殊的非负矩阵,即随机矩阵与双随机矩阵的知识及其应用.

n 阶非负矩阵的某些明显性质可由如下事实推出:它们组成一个凸锥,即若 A 与 B 属于 n 阶非负矩阵集合 $\mathscr{N}_n$,则 $\alpha A+\beta B\in\mathscr{N}_n$, $\forall\alpha,\beta\in\mathbb{R}_+$(非负实数的集合),并且,$\mathscr{N}_n$ 是一个偏序集(对于序关系"$\geqslant$").

记

$$\mathbb{R}_+^n=\{\boldsymbol{x}=(x_1,\cdots,x_n)^{\mathrm{T}}\in\mathbb{R}^n:\text{所有 } x_j\geqslant 0\};$$

$$\boldsymbol{H}^n=\mathbb{R}_+^n\setminus\{\boldsymbol{0}\};\boldsymbol{E}^n=\{\boldsymbol{x}=(x_1\cdots,x_n)^{\mathrm{T}}\in\boldsymbol{H}^n:\sum_{j=1}^n x_j=1\}.$$

显然,$A\in\mathscr{N}_n$ 当且仅当 $A\mathbb{R}_+^n\subset\mathbb{R}_+^n$. 因此,$A\in\mathscr{N}_n$ 与 $\boldsymbol{x}\geqslant\boldsymbol{y}(\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n)$ 蕴涵 $A\boldsymbol{x}\geqslant A\boldsymbol{y}$. 若 $A,B,C\in M_n(\mathbb{R})$,且 $A\geqslant O$ 与 $B\geqslant C$,则 $AB\geqslant AC$. 并且,$A\in M_n(\mathbb{R})$ 为正矩阵,当且仅当 $A\boldsymbol{H}^n\subset\mathrm{int}\mathbb{R}_+^n$,这里,$\mathrm{int}\mathbb{R}_+^n$ 表示$\mathbb{R}_+^n$ 的内部,即

$$\mathrm{int}\mathbb{R}_+^n=\{\boldsymbol{x}=(x_1,\cdots,x_n)^{\mathrm{T}}>\boldsymbol{0}\}.$$

因此,若 $A>O$ 且 $\boldsymbol{x}\geqslant\boldsymbol{y}$ 但 $\boldsymbol{x}\neq\boldsymbol{y}$,则 $A\boldsymbol{x}>A\boldsymbol{y}$. 若 $A,B\in\boldsymbol{M}_n(\mathbb{R})$,且 $A\geqslant O,B>O$ 与 $AB=O$,则有 $A=O$,容易看出,$A\in\boldsymbol{M}_n(\mathbb{R})$ 非负(正)当且仅当 A^{T} 非负(正);A 不可约当且仅当 A^{T} 不可约,或等价地,当且仅当 $|A|\equiv(|a_{ij}|)\in\boldsymbol{M}_n(\mathbb{R})$ 非负不可约.

7.1 非负不可约矩阵的 Perron-Frobenius 理论

7.1.1 最基本的结果

推演 Perron-Frobenius 理论基本定理的方法有多种(见文献[2]或文献[10]),我们这里主要采用 Wielandt[14] 的办法. 为此先引入 **Collatz-Wielandt 函数**,利用它的性质可以比较自然地得出一些基本结果.

定义 1　设 $A\in\mathcal{N}_n$ 不可约. 如下函数 $F:\boldsymbol{H}^n\to\mathbb{R}_+$

$$F(\boldsymbol{x})=\min_{x_j>0}\frac{(A\boldsymbol{x})_j}{x_j},\qquad \forall\,\boldsymbol{x}\in\boldsymbol{H}^n \tag{1}$$

叫作对应于 A 的 **Collatz-Wielandt 函数**,简称为 A 的 C-W 函数.

例如,对于三阶非负不可约矩阵

$$A=\begin{bmatrix}2&2&1\\2&2&1\\0&2&1\end{bmatrix}, \tag{2}$$

若 $\boldsymbol{x}=(1,0,0)^{\mathrm{T}}\in\boldsymbol{H}^3$,则有 $F(\boldsymbol{x})=(A\boldsymbol{x})_1=2$;若 $\boldsymbol{x}_\varepsilon=(1-\varepsilon,0,\varepsilon)^{\mathrm{T}}(0<\varepsilon<1)$,则有 $F(\boldsymbol{x}_\varepsilon)=\min\{(2-\varepsilon)/(1-\varepsilon),1\}=1$.

显然,(1)式中 F 只依赖于 A,且 $0\leqslant F(\boldsymbol{x})<\infty$,$F(\boldsymbol{x})\boldsymbol{x}\leqslant A\boldsymbol{x}$,$\forall\,\boldsymbol{x}\in\boldsymbol{H}^n$. C-W 函数的另外一些初等性质罗列如下.

命题 2　设 $A\in\mathcal{N}_n$ 不可约,F 为 A 的 C-W 函数. 则有

(1) F 为(零次)齐次与有界的.

(2) 若 ρ 为满足 $A\boldsymbol{x}\geqslant\rho\boldsymbol{x}$,$\boldsymbol{x}\in\boldsymbol{H}^n$ 的最大实数,则 $\rho=F(\boldsymbol{x})$,亦即 $F(\boldsymbol{x})=\max\{r\in\mathbb{R}:A\boldsymbol{x}\geqslant r\boldsymbol{x}\}$,$\forall\,\boldsymbol{x}\in\boldsymbol{H}^n$.

(3) 若 $\boldsymbol{x}\in\boldsymbol{H}^n$ 且 $\boldsymbol{y}=(I+A)^{n-1}\boldsymbol{x}$,则 $F(\boldsymbol{x})\leqslant F(\boldsymbol{y})$.

证明　(1) F 显然为零次齐次的,因为有 $F(t\boldsymbol{x})=F(\boldsymbol{x})$,$\forall\,t>0$ 与 $\forall\,\boldsymbol{x}\in\boldsymbol{H}^n$. 由于 F 以 0 为下界,故只要证明 F 有上界即可. 事实上,我们有 $F(\boldsymbol{x})x_j\leqslant(A\boldsymbol{x})_j$,$\forall\,\boldsymbol{x}\in\boldsymbol{H}^n$,$j=1,\cdots,n$,相加之得出,

$$F(\boldsymbol{x})\,\|\boldsymbol{x}\|_1\leqslant\|A\boldsymbol{x}\|_1\leqslant\|A\|_1\,\|\boldsymbol{x}\|_1,$$

因而有

$$F(\boldsymbol{x})\leqslant\|A\|_1,\ \forall\,\boldsymbol{x}\in\boldsymbol{H}^n. \tag{3}$$

(2) 由定义 1 知道,$A\boldsymbol{x}\geqslant F(\boldsymbol{x})\boldsymbol{x}$,且因 $\boldsymbol{x}\in\boldsymbol{H}^n$,所以有分量 $x_k>0$ 使得 $A\boldsymbol{x}-F(\boldsymbol{x})\boldsymbol{x}$ 的第 k 个分量为零:$(A\boldsymbol{x})_k-F(\boldsymbol{x})x_k=0$. 于是,若 $F(\boldsymbol{x})<\alpha$,则 $A\boldsymbol{x}-\alpha\boldsymbol{x}$ 必有一个负的分量,因而 α 不满足 $A\boldsymbol{x}-\alpha\boldsymbol{x}\geqslant\boldsymbol{0}$.

(3) 若 $\boldsymbol{x}\in\boldsymbol{H}^n$,则有 $F(\boldsymbol{x})\boldsymbol{x}\leqslant A\boldsymbol{x}$. 用 $(I+A)^{n-1}\in\mathcal{N}_n$ 左乘前式得出

$$F(\boldsymbol{x})(I+A)^{n-1}\boldsymbol{x}\leqslant(I+A)^{n-1}A\boldsymbol{x}=A(I+A)^{n-1}\boldsymbol{x},$$

$$F(\boldsymbol{x})\boldsymbol{y}\leqslant A\boldsymbol{y}.$$

因为 $\boldsymbol{y}\in\boldsymbol{H}^n$,所以由(2)得出,$F(\boldsymbol{x})\leqslant F(\boldsymbol{y})$.　□

下面结果提供非负不可约矩阵的一种特征.

定理 3　设 $A\in\mathcal{N}_n$,$n\geqslant2$,则 A 不可约当且仅当

$$(I+A)^{n-1}>O.$$

证明　设 A 非负不可约,要证 $(I+A)^{n-1}\boldsymbol{x}>\boldsymbol{0}$,$\forall\,\boldsymbol{x}\in\boldsymbol{H}^n$. 令 $\boldsymbol{y}=(I+A)\,\boldsymbol{x}\in\boldsymbol{H}^n$. 若 $\boldsymbol{y}>\boldsymbol{0}$,则更有 $(I+A)^{n-1}\boldsymbol{x}>\boldsymbol{0}$;若 $\boldsymbol{y}$ 有 m 个分量 $y_i=0$,$1\leqslant m<n$,则因 $\boldsymbol{0}\leqslant$

$x \leqslant y$,所以 x 的相应分量 $x_i=0$. 此时存在置换矩阵 P,使得 $P\boldsymbol{y}=\begin{bmatrix}\boldsymbol{\alpha}\\ \boldsymbol{0}\end{bmatrix}$与 $P\boldsymbol{x}=\begin{bmatrix}\boldsymbol{\beta}\\ \boldsymbol{0}\end{bmatrix}$,

其中,$\boldsymbol{\alpha}\in\mathbb{R}^{n-m}$为正向量,$\boldsymbol{\beta}\in\mathbb{R}^{n-m}$为非负向量. 现设

$$PAP^{\mathrm{T}}=\begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix},$$

其中 A_{11} 为 $n-m$ 阶矩阵. 从 $\boldsymbol{y}=(I+A)\boldsymbol{x}$ 得出

$$\begin{bmatrix}\boldsymbol{\alpha}\\ \boldsymbol{0}\end{bmatrix}=\begin{bmatrix}\boldsymbol{\beta}\\ \boldsymbol{0}\end{bmatrix}+\begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix}\begin{bmatrix}\boldsymbol{\beta}\\ \boldsymbol{0}\end{bmatrix},$$

因而 $A_{21}\boldsymbol{\beta}=\boldsymbol{0}$. 因 $A\in\mathcal{N}_n$ 不可约,我们有 $A_{21}\neq O$,于是 $\boldsymbol{\beta}\geqslant\boldsymbol{0}$ 有零分量. 此表明 $\boldsymbol{y}$ 的零分量个数严格小于 $\boldsymbol{x}$ 的零分量个数. 但 $\boldsymbol{x}$ 至多有 $n-1$ 个零分量,故有$(I+A)^{n-1}\boldsymbol{x}>\boldsymbol{0}$. 反之,设 $A\in\mathcal{N}_n$ 满足$(I+A)^{n-1}>O$. 假若 A 可约,按定义有置换矩阵 P 使得第六章 6.1.2 小节(1)式成立,于是,

$$P^{\mathrm{T}}(I+A)^{n-1}P=(P^{\mathrm{T}}(I+A)P)^{n-1}=\begin{bmatrix}I+B & C\\ O & I+D\end{bmatrix}^{n-1}=\begin{bmatrix}* & *\\ O & *\end{bmatrix},$$

即有$(I+A)^{n-1}\not>O$. 与假设矛盾. □

我们指出从上述证明顺便得到,可约矩阵的任意乘幂仍为可约矩阵. 因此,若 $A\in\boldsymbol{M}_n(\mathbb{R})$且对某正整数 p,$A^p>O$,则 A 一定不可约. 此外,显然地,由定理 3 我们还得到,$A\in\boldsymbol{M}_n(\mathbb{C})(n\geqslant2)$为不可约矩阵当且仅当$(I+|A|)^{n-1}>O$,这里$|A|=(|a_{ij}|)\in\mathcal{N}_n$(见第六章习题 6.2.2 题 4(1)).

命题 4　设 $A\in\mathcal{N}_n$ 不可约,且 F 为它的 C-W 函数. 则存在 $\boldsymbol{u}_0>\boldsymbol{0}$ 使得

$$F(\boldsymbol{u}_0)=\max_{\boldsymbol{x}\in\boldsymbol{E}^n}F(\boldsymbol{x})=\max_{\boldsymbol{x}\in\boldsymbol{H}^n}F(\boldsymbol{x}).\tag{4}$$

证明　当 $n=1$ 时,取 $\boldsymbol{u}_0=(1)$,则(4) 式成立. 对 $n\geqslant2$,令 $\widetilde{\boldsymbol{H}}^n=\{\boldsymbol{u}\in\boldsymbol{H}^n:\boldsymbol{u}=\boldsymbol{y}/\sum_{i=1}^n y_i,\boldsymbol{y}=((I+A)^{n-1}\boldsymbol{x},\ \forall\boldsymbol{x}\in\boldsymbol{E}^n)\}\subset\boldsymbol{H}^n$. 按定理 3,$\widetilde{\boldsymbol{H}}^n$ 中所有向量 $\boldsymbol{u}>\boldsymbol{0}$. 容易验证,$\widetilde{\boldsymbol{H}}^n\subset\boldsymbol{E}^n$ 为有界闭集,因而 $\widetilde{\boldsymbol{H}}^n$ 上连续函数 F 在某个 $\boldsymbol{u}_0\in\widetilde{\boldsymbol{H}}^n$ 处达到最大值. 按命题 2 中(1) 与(3),对任意 $\boldsymbol{x}\in\boldsymbol{H}^n$ 或 $\boldsymbol{x}\in\boldsymbol{E}^n$,令 $\boldsymbol{y}=(I+A)^{n-1}\boldsymbol{x}$,则有

$F(\boldsymbol{x})\leqslant F(\boldsymbol{y})=F\left(\boldsymbol{y}/\sum_{i=1}^n y_i\right)\leqslant F(\boldsymbol{u}_0)$. 因此,(4) 式成立. □

我们指出,在上述证明中,不能用 $\boldsymbol{E}^n$ 代替 $\widetilde{\boldsymbol{H}}^n$,这是因为 F 一般不是 $\boldsymbol{E}^n$ 上的连续函数. 一个简单的反例见(2)式中矩阵,这时 F 在 $\boldsymbol{x}_0=(1,0,0)^{\mathrm{T}}\in\boldsymbol{E}^3$ 处不连续.

有了这些预备知识,现在可提出并证明非负不可约矩阵的 Perron-Frobenius 理论的最基本(也是最重要)的结果.

定理 5　设 $A\in\mathcal{N}_n$ 不可约. 则 A 有一个正特征值 r 等于 A 的谱半径 $\rho(A)$,且对应于 $\rho(A)$,A 有一个正的特征向量.

证明 按命题 4 有 $\mathbf{0}<\boldsymbol{u}_0\in\boldsymbol{H}^n$ 满足(4)式. 现令 $r=F(\boldsymbol{u}_0)$ 与 $\boldsymbol{x}=(1,\cdots,1)^{\mathrm{T}}\in\boldsymbol{H}^n$，则 $F(\boldsymbol{x})=\min\limits_{i}\sum\limits_{j=1}^{n}a_{ij}>0$，于是，$r=F(\boldsymbol{u}_0)\geqslant F(\boldsymbol{x})>0$. 另一方面，按定义 1，$r\boldsymbol{u}_0\leqslant A\boldsymbol{u}_0$. 我们证明：$A\boldsymbol{u}_0=r\boldsymbol{u}_0$. 如不然则有 $A\boldsymbol{u}_0-r\boldsymbol{u}_0\in\boldsymbol{H}^n$. 令 $\boldsymbol{y}=(I+A)^{n-1}\boldsymbol{u}_0$. 则 $\boldsymbol{y}>\mathbf{0}$，且 $(I+A)^{n-1}(A\boldsymbol{u}_0-r\boldsymbol{u}_0)=A\boldsymbol{y}-r\boldsymbol{y}>\mathbf{0}$，因而 $r<F(\boldsymbol{y})$，与 $r=F(\boldsymbol{u}_0)=\max\limits_{\boldsymbol{x}\in\boldsymbol{H}^n}F(\boldsymbol{x})$ 矛盾. 余下只要证明：$r=\rho(A)$. 设 $\lambda\in\sigma(A)$ 与 $A\boldsymbol{u}=\lambda\boldsymbol{u},\boldsymbol{u}\neq\mathbf{0}$. 则有 $|\lambda||\boldsymbol{u}|\leqslant A|\boldsymbol{u}|$，因而按命题 2 中(2)以及 r 的定义，$|\lambda|\leqslant F(|\boldsymbol{u}|)|\leqslant r\in\sigma(A)$，即有 $r=\rho(A)$. □

在许多文献中，将不可约矩阵 $A\in\mathcal{N}_n$ 的特征值 $\rho(A)$ 叫作 A 的 **Perron 根**，其对应的正特征向量叫作 A 的 **Perron 向量**. 对于非负不可约矩阵 A 的 Perron 根与 Perron 向量，我们有

定理 6 设 $A\in\mathcal{N}_n$ 不可约.

(1)若 $A\boldsymbol{x}=\rho(A)\boldsymbol{x}$ 与 $\boldsymbol{x}\in\boldsymbol{H}^n$，则 $\boldsymbol{x}>\mathbf{0}$. 即 $\boldsymbol{x}$ 为 A 的 Perron 向量.

(2)若 A 有非负特征向量 $\boldsymbol{x}$ 对应于 $\lambda\in\sigma(A)$，则 $\lambda=\rho(A)$，即 λ 必为 A 的 Perron 根，且 $\boldsymbol{x}>\mathbf{0}$，即 $\boldsymbol{x}$ 必为 A 的 Perron 向量.

证明 当 $n=1$ 时定理结论显然成立. 当 $n\geqslant2$ 时，(1)中假定条件蕴涵 $(I+A)^{n-1}\boldsymbol{x}=(1+\rho(A))^{n-1}\boldsymbol{x}$ 与 $\boldsymbol{x}\in\boldsymbol{H}^n$. 按定理 3，$(I+A)^{n-1}\boldsymbol{x}>\mathbf{0}$，于是 $(1+\rho(A))^{n-1}\boldsymbol{x}>\mathbf{0}$ 蕴涵 $\boldsymbol{x}>\mathbf{0}$. 若 $A\boldsymbol{x}=\lambda\boldsymbol{x}$ 与 $\boldsymbol{x}\in\boldsymbol{H}^n$，将定理 5 用于 A^{T}，我们有 $\boldsymbol{z}>\mathbf{0}$，使得 $A^{\mathrm{T}}\boldsymbol{z}=\rho(A)\boldsymbol{z}$，那么，

$$\lambda\langle\boldsymbol{x},\boldsymbol{z}\rangle=\langle A\boldsymbol{x},\boldsymbol{z}\rangle=\langle\boldsymbol{x},A^{\mathrm{T}}\boldsymbol{z}\rangle=\rho(A)\langle\boldsymbol{x},\boldsymbol{z}\rangle,$$

式中，$\langle,\rangle$ 表示 $\mathbb{R}^n$ 中标准内积. 但因 $x\in\boldsymbol{H}^n$ 与 $\boldsymbol{z}>\mathbf{0}$ 蕴涵 $\langle\boldsymbol{x},\boldsymbol{z}\rangle>0$，所以上式推出 $\lambda=\rho(A)$. 再由(1)便知，$\boldsymbol{x}$ 为 A 的 Perron 向量. □

现在我们来改进定理 5 的基本结果. 先讨论 Wielandt 的一个结果，它是后面许多结论的证明的有力工具.

定理 7(Wielandt) 设 $A\in\mathcal{N}_n$ 不可约，$B\in\boldsymbol{M}_n(\mathbb{C})$.

(1) 若 $|B|\leqslant A$，则对任意 $\beta\in\sigma(B)$，

$$|\beta|\leqslant\rho(A),\tag{5}$$

因而 $\rho(B)\leqslant\rho(A)$.

(2) (5)式中等式成立当且仅当 B 有形式

$$B=\mathrm{e}^{\mathrm{i}\varphi}DAD^{-1},\tag{6}$$

其中，$\mathrm{e}^{\mathrm{i}\varphi}=\beta/\rho(A)$，$0\leqslant\varphi<2\pi$，与 $|D|=I$.

证明 (1)设 $B\boldsymbol{y}=\beta\boldsymbol{y}$，$\boldsymbol{y}\neq\mathbf{0}$. 则 $|\boldsymbol{y}|\in\boldsymbol{H}^n$，并且，

$$|\beta||\boldsymbol{y}|\leqslant|B||\boldsymbol{y}|\leqslant A|\boldsymbol{y}|.\tag{7}$$

按照命题 2 中(2)与定理 5 的证明，(7)式蕴涵

$$|\beta|\leqslant F(|\boldsymbol{y}|)\leqslant\rho(A).\tag{8}$$

(2) 若$|\beta|=\rho(A)$，即$\beta=e^{i\varphi}\rho(A)$对某实数$\varphi\in[0,2\pi)$成立，则(7)式推出，$\rho(A)|\boldsymbol{y}|\leqslant A|\boldsymbol{y}|$. 应用在定理5证明中的类似推证以及定理6中(1)，我们有

$$A|\boldsymbol{y}|=\rho(A)|\boldsymbol{y}|,\qquad |\boldsymbol{y}|>\boldsymbol{0}.$$

因此，(7)式变为等式：

$$|\beta||\boldsymbol{y}|=|B||\boldsymbol{y}|=A|\boldsymbol{y}|. \tag{9}$$

但$|B|\leqslant A$与$|\boldsymbol{y}|>\boldsymbol{0}$，故有$|B|=A$. 令$D=\mathrm{diag}[y_1/|y_1|,\cdots,y_n/|y_n|]$与$G=(g_{ij})=e^{-i\varphi}D^{-1}BD$，则显然有$\boldsymbol{y}=D|\boldsymbol{y}|$，$|D|=I$与$|G|=|B|=A$. 于是，从$B\boldsymbol{y}=\beta\boldsymbol{y}$，$\boldsymbol{y}\neq\boldsymbol{0}$推出

$$BD|\boldsymbol{y}|=\beta D|\boldsymbol{y}|=\rho(A)e^{i\varphi}D|\boldsymbol{y}|,$$

进而$G|\boldsymbol{y}|=A|\boldsymbol{y}|=|G||\boldsymbol{y}|$，即有$\sum_{j=1}^{n}(|g_{ij}|-g_{ij})|y_j|=0, i=1,\cdots,n$. 由于$|y_j|>0$，$\forall i$，故由前一式子推得$G=|G|=A$，此相当于(6)式. 反之，若$B$有形式(6)，则$|B|=A$，且$B$与$e^{i\varphi}A$为相似矩阵，因而$B$必有一特征值$\beta$满足$\beta=e^{i\varphi}\rho(A)$. □

上述定理有如下两个有用的推论.

推论8　设$A\in\mathcal{N}_n$不可约，$n\geqslant2$. 若C为A的任意主子阵且$C\neq A$，则$\rho(C)<\rho(A)$.

证明　不妨设C为A的第m个前主子阵$A[(1,\cdots,m)]$，$1\leqslant m<n$，且令$B=\mathrm{diag}[C,O]\in\mathcal{N}_n$. 则$O\leqslant B\leqslant A$，$\rho(C)=\rho(B)$. 但因$B=|B|\neq A$，所以定理7断定，$\rho(C)=\rho(B)<\rho(A)$. □

应用$O\leqslant A\leqslant B$与A不可约蕴涵$B\in\mathcal{N}_n$不可约的事实，从定理7直接得到下列结果.

推论9　设$A\in\mathcal{N}_n$不可约. 则当A的任一元素增大时，$\rho(A)$也随之增大.

对于非负不可约矩阵A，由定理5知道，$\rho(A)$为它的正特征值. 下面结论进一步指出，$\rho(A)$为A的单重(指代数重数)特征值，即$\rho(A)$为$\det(\lambda I-A)=0$的单重根.

定理10　设$A\in\mathcal{N}_n$不可约. 则$\rho(A)$为A的单重特征值，因而对应于$\rho(A)$的A的特征空间$\mathrm{Ker}(A-\rho(A)I)$为一维的，即$\rho(A)$的几何重数也等于1.

证明　显然只要考虑$n\geqslant2$的情形. 为证$\rho(A)$的单重性，只要验明$\Delta(\lambda)\equiv\det(\lambda I-A)$满足$\Delta'(\rho(A))\neq0$. 应用$\lambda$-矩阵的行列式求导数公式有

$$\Delta'(\lambda)=\sum_{j=1}^{n}\det((\lambda I-A)(j|j)), \tag{10}$$

式中$B(j|j)$表示方阵B去掉第j行与第j列余下的主子阵. 按推论8，$\det((\lambda I-A)(j|j))\neq0$，$\forall\lambda\geqslant\rho(A)$. 但当$\lambda$足够大时，$\det((\lambda I-A)(j|j))>0$，因而此式对所有$\lambda\geqslant\rho(A)$成立，$j=1,\cdots,n$. 代入(10)式便得$\Delta'(\lambda)>0$，$\forall\lambda\geqslant\rho(A)$. 容易看出，本

定理的后一半结果为前一部分的直接推论. □

由上述定理的证明我们可以顺便地得到如下事实.

推论 11　设 $A\in\mathcal{N}_n$ 不可约且令 $B(\lambda)\equiv\text{adj}(\lambda I-A)$. 则 $B(\rho(A))>O$.

证明　令 $\Delta(\lambda)\equiv\det(\lambda I-A)$ 与 $r=\rho(A)$. 则有 $(\lambda I-A)B(\lambda)=\Delta(\lambda)I$,因而从 $\Delta'(r)=\text{tr}B(r)>0$ 与 $B(r)$ 所有对角元素皆为正数的事实(见定理 10 证明)得出, $B(r)$ 的每一列向量为对应于 r 的 A 的特征向量. 再按定理 10,这些列向量一定为正向量(因有一个元素为正数),即有 $B(r)>O$. □

对于非负不可约矩阵 A 的所有与 $\rho(A)$ 同模的特征值,下一个推论保证它们都为单重的(我们顺便指出,这类单重的性质对非负不可约矩阵的特征值扰动分析十分有利).

推论 12　设 $A\in\mathcal{N}_n$ 不可约. 则模为 $\rho(A)$ 的 A 的特征值皆为单重的.

证明　设 $\beta\in\sigma(A)$ 满足 $|\beta|=\rho(A)$. 则有 $\varphi\in[0,2\pi)$ 使得 $\beta=e^{i\varphi}\rho(A)$. 应用定理 7(取 $B=A$),我们有

$$A = e^{i\varphi}DAD^{-1}, \qquad |D| = I, \tag{11}$$

因而 A 与 $e^{i\varphi}A$ 相似. 按定理 10, $\rho(A)$ 的单重性这时蕴涵 $\beta=e^{i\varphi}\rho(A)$ 的单重性. □

在提出非负不可约矩阵的 Perron-Frobenius 理论第一部分结论之前,我们应用定理 6 与定理 10 给出非负不可约矩阵的另外几种特征.

定理 13　设 $A\in\mathcal{N}_n, n\geqslant 2$,则下列每个命题都等价于"$A$ 为不可约矩阵":

(1) A 的所有特征向量不在 $\mathbb{R}_+^n$ 的边界面上.

(2) 如不计常数因子, A 只有一个非负的特征向量,且它为正的.

(3) 若 $\alpha\boldsymbol{x}\geqslant A\boldsymbol{x}$,其中 $\alpha\in\mathbb{R}$ 与 $\boldsymbol{x}\in\boldsymbol{H}^n$,则 $\boldsymbol{x}>\boldsymbol{0}$.

证明　只证 $A\in\mathcal{N}_n$ 不可约与(1)以及(2)之间的等价性,与(3)的等价性的证明留给读者(见习题 7). $A\in\mathcal{N}_n$ 不可约蕴涵(2)是显然的,因为它是定理 6 与定理 10 的直接推论. (2)显然蕴涵(1). 余下证明:(1)蕴涵 A 不可约. 设 $A\in\mathcal{N}_n(n\geqslant 2)$ 且(1)成立. 倘若 A 可约,则有置换矩阵 P 使得第六章 6.1.2 小节(1)式成立. 不妨设其中的 $B\geqslant O$ 为一阶零矩阵或不可约的(否则对 B 可以继续约化使得这个条件成立). 应用定理 5,这时有 $\boldsymbol{x}>\boldsymbol{0}$ 使得 $B\boldsymbol{x}=\rho(B)\boldsymbol{x}$,因而 A 有特征向量 $\boldsymbol{y}=P(\boldsymbol{x},\boldsymbol{0})^{\mathrm{T}}$,使得 $A\boldsymbol{y}=\rho(B)\boldsymbol{y}$,此与(1)矛盾. 因此 $A\in\mathcal{N}_n$ 必定不可约. □

现在我们提出非负不可约矩阵的 Perron-Frobenius 理论的第一部分结果.

定理 14　(1)若 $A\in\mathcal{N}_n$ 为正矩阵,则 $\rho(A)$ 为 A 的单重正特征值,且 A 的其他特征值的模严格小于 $\rho(A)$. 对应于 $\rho(A)$,有 $\boldsymbol{x}>\boldsymbol{0}$ 使得 $A\boldsymbol{x}=\rho(A)\boldsymbol{x}$.

(2)若 $A\in\mathcal{N}_n$ 不可约,则 $\rho(A)$ 为 A 的单重正特征值,且模为 $\rho(A)$ 的所有 A 的特征值也是单重的. 对应于 $\rho(A)$, A 有一个正特征向量,且 A 的任何非负特征向量与此正特征向量线性相关.

证明　由前面讨论知道,(2)成立,因而对(1)中结果只需证明:若 $A\in\mathcal{N}_n$ 为正

矩阵,$\lambda\in\sigma(A)$且$\lambda\neq\rho(A)$,则$|\lambda|<\rho(A)$. 现设$Ay=\lambda y$,$y\neq\mathbf{0}$且$|\lambda|=\rho(A)$. 应用定理7证明(取$B=A$与$\beta=\lambda$),我们有$\rho(A)|\boldsymbol{y}|=A|\boldsymbol{y}|$与$|\boldsymbol{y}|>\mathbf{0}$,且$|A\boldsymbol{y}|=A|\boldsymbol{y}|$,因而特别有

$$\left|\sum_{j=1}^{n}a_{1j}y_j\right|=\sum_{j=1}^{n}a_{1j}\,|\,y_j\,|. \tag{12}$$

由于$a_{11},a_{12},\cdots,a_{1n}$均为正数,故若将$y_j$写成极分解形式$y_j=|y_j|e^{i\theta_j}(1\leqslant j\leqslant n)$,则有

$$\left|\sum_{j=1}^{n}a_{1j}\,|\,y_j\,|\,e^{i\theta_j}\right|=\sum_{j=1}^{n}a_{1j}\,|\,y_j\,|.$$

这表明所有复数y_j同向,即存在一个与j无关的θ使得$y_j=|y_j|e^{i\theta_j}$,$j=1,\cdots,n$,或等价地有$\boldsymbol{y}=e^{i\theta_j}|\boldsymbol{y}|$. 因此,$\boldsymbol{y}$也是对应于$\rho(A)$的$A$的特征向量,于是,$(\rho(A)-\lambda)\boldsymbol{y}=A\boldsymbol{y}-A\boldsymbol{y}=\mathbf{0}$,这必定有$\lambda=\rho(A)$. □

定理14(1)始于Perron,故亦称之为**Perron定理**. 应用它可以得到矩阵的一种排除定理.

推论15(Ky Fan)　设$A=(a_{ij})\in\boldsymbol{M}_n(\mathbb{C})$与$B=(b_{ij})\in\mathcal{N}_n$满足$|A|\leqslant B$. 则有

$$\sigma(A)\subset\bigcup_{i=1}^{n}\{z\in\mathbb{C}:\,|\,z-a_{ii}\,|\leqslant\rho(B)-b_{ii}\}. \tag{13}$$

证明　不妨设$B>0$,否则可考虑$B_\varepsilon=(b_{ij}+\varepsilon)$,$\varepsilon>0$,这时有$B_\varepsilon>|A|$与$\rho(B_\varepsilon)-(b_{ii}+\varepsilon)\to\rho(B)-b_{ii}(\varepsilon\to0)$. 根据Perron定理,有$\boldsymbol{x}>\mathbf{0}$使得$B\boldsymbol{x}=\rho(B)\boldsymbol{x}$,因而

$$\sum_{j\neq i}|\,a_{ij}\,|\,x_j\leqslant\sum_{j\neq i}b_{ij}x_j=\rho(B)x_i-b_{ii}x_i,\qquad\forall i,$$

由此推出

$$\frac{1}{x_i}\sum_{j\neq i}|\,a_{ij}\,|\,x_j\leqslant\rho(B)-b_{ii},\qquad\forall i.$$

再应用第六章6.1.1小节(8)式即可. □

对于非负不可约矩阵,定理14连同定理6与推论9可以给出谱半径的估计界限,它在理论上尤其在矩阵迭代分析中有着重要作用.

引理16　设$A\in\mathcal{N}_n$不可约. 则有

$$\sum_{j=1}^{n}a_{ij}=\rho(A),\qquad i=1,\cdots,n, \tag{14}$$

或者

$$\min_{1\leqslant i\leqslant n}\sum_{j=1}^{n}a_{ij}<\rho(A)<\max_{1\leqslant i\leqslant n}\sum_{j=1}^{n}a_{ij}. \tag{15}$$

证明　若A的每行元素之和均等于σ,取$\xi=(1,\cdots,1)^{\mathrm{T}}\in\boldsymbol{H}^n$,则$A\xi=\sigma\xi$,因而按定理6中(2)有$\rho(A)=\sigma$. 若$A$的各行和不全相等,则可以用减小(或增加)$A$的某些正元素的方法得到一个非负不可约矩阵$B=(b_{ij})$(或$C=(c_{ij})$)使得对所有$i\in\mathcal{N}$,都有

$$\sum_{j=1}^{n} b_{ij} = \alpha \equiv \min_{i\in\mathcal{N}} \sum_{j=1}^{n} a_{ij},$$

$$\sum_{j=1}^{n} c_{ij} = \beta \equiv \max_{i\in\mathcal{N}} \sum_{j=1}^{n} a_{ij}.$$

此时,$O\leqslant B\leqslant A\leqslant C$ 且 $B\neq A$ 与 $C\neq A$. 应用刚才证明的结果有 $\rho(B)=\alpha$ 与 $\rho(C)=\beta$. 再引用推论 9 便得 $\alpha<\rho(A)<\beta$,此即(15)式. □

定理 17 设 $A\in\mathcal{N}_n$ 不可约. 则对任意给定正向量 $\boldsymbol{x}=(x_1,\cdots,x_n)^{\mathrm{T}}$,我们有

$$\frac{1}{x_i}\sum_{j=1}^{n} a_{ij}x_j = \rho(A), \qquad i=1,\cdots,n, \tag{16}$$

或者

$$\min_{1\leqslant i\leqslant n}\left(\frac{1}{x_i}\sum_{j=1}^{n} a_{ij}x_j\right) < \rho(A) < \max_{1\leqslant i\leqslant n}\left(\frac{1}{x_i}\sum_{j=1}^{n} a_{ij}x_j\right). \tag{17}$$

此外,

$$\max_{x>0}\min_{1\leqslant i\leqslant n}\left(\frac{1}{x_i}\sum_{j=1}^{n} a_{ij}x_j\right) = \rho(A) = \min_{x>0}\max_{1\leqslant i\leqslant n}\left(\frac{1}{x_i}\sum_{j=1}^{n} a_{ij}x_j\right). \tag{18}$$

证明 给定 $\boldsymbol{x}=(x_1,\cdots,x_n)^{\mathrm{T}}>\boldsymbol{0}$,令 $D=\mathrm{diag}[x_1,\cdots,x_n]$. 将引理 16 用于非负不可约矩阵 $B=D^{-1}AD$,直接推出本定理的第一部分. 由(16)式与(17)式看出,当(18)式中"="换成"$\leqslant$"时,相应不等式成立. 再选取 $\boldsymbol{x}$ 为对应于 $\rho(A)$ 的 A 的正特征向量,便知道(18)式中两个等式同时成立. □

上述定理(18)式中两个关于 $\rho(A)$ 的等式有时也分别地称为 **max-min(极大极小)判定与 min-max(极小极大)判定**. 显然,对于列和,引理 16 与定理 17 也有类似的结论,它也提供 $\rho(A)$ 位置的界限.

例如对 $A=(a_{ij})\in\mathcal{N}_n$, $a_{ij}=1$, $\forall\, i,j\in\mathcal{N}$, 由引理 16 知道,$\rho(A)=n$(单重特征值),而其他特征值按 Perron 定理其模均小于 n(实际上它们皆为零,因为 A 实对称正半定且 $\mathrm{rank}A=1$).

定理 17 有如下推论,其证明留给读者(见习题 8).

推论 18 设 $A\in\mathcal{N}_n$ 不可约. 则

$$\alpha\boldsymbol{x}\leqslant A\boldsymbol{x}\ \text{对某个}\ \boldsymbol{x}>\boldsymbol{0}\ \text{与}\ \alpha\in\mathbb{R} \tag{19}$$

成立蕴涵 $\alpha\leqslant\rho(A)$,并且

$$A\boldsymbol{x}\leqslant\beta\boldsymbol{x}\ \text{对某个}\ \boldsymbol{x}>\boldsymbol{0}\ \text{与}\ \beta\in\mathbb{R} \tag{20}$$

成立蕴涵 $\rho(A)\leqslant\beta$. 此外,若(19)式中还有 $\alpha\boldsymbol{x}\neq A\boldsymbol{x}$,则 $\alpha<\rho(A)$;若(20)式中还有 $A\boldsymbol{x}\neq\beta\boldsymbol{x}$,则 $\rho(A)<\beta$.

习题 7.1.1

1. 设

$$A=\begin{bmatrix}1 & 3 & 4\\ 2 & 1 & 1\\ 3 & 4 & 5\end{bmatrix},$$

且 F 为 A 的 C-W 函数. 对 $\boldsymbol{x}=(1,1,1)^{\mathrm{T}},(1,0,2)^{\mathrm{T}},(2,1,3)^{\mathrm{T}}$, 求 $F(\boldsymbol{x})$, 并求出 $\rho(A)$ 的一个下界.

2. 设 $A\in\mathcal{N}_n$ 不可约且 F 为 A 的 C-W 函数. 试寻找出一个有关 A 的充分与必要条件使得 F 在 $\boldsymbol{E}^n$ 上连续.

3. 设 $n\geqslant 2$ 与 $\mathcal{M}\subset\mathcal{N}=\{1,\cdots,n\}$. 称 F_μ 为 $\mathbb{R}_+^n$ 的一个面, 假如 $F_\mu=\{\boldsymbol{x}=(x_1,\cdots,x_n)^{\mathrm{T}}\in\mathbb{R}_+^n: x_i=0,\forall i\notin\mathcal{M}\}$. 试证: $A\in\mathcal{N}_n$ 不可约当且仅当满足 $AF_\mu\subset F_\mu$ 的 $\mathbb{R}_+^n$ 的面只是 $F_\mu=\{\boldsymbol{0}\}$ 或 $F_\mu=\mathbb{R}_+^n$. 并且应用这个结果证明如下 $n(\geqslant 2)$ 阶置换矩阵

$$P=\begin{bmatrix}0 & 1 & 0 & 0 & \cdots & 0 & 0\\ 0 & 0 & 1 & 0 & \cdots & 0 & 0\\ 0 & 0 & 0 & 1 & \cdots & 0 & 0\\ \vdots & & & \ddots & \ddots & & \vdots\\ 0 & 0 & 0 & 0 & \cdots & 0 & 1\\ 1 & 0 & 0 & 0 & \cdots & 0 & 0\end{bmatrix}$$

为非负不可约的.

4. 试证: 若 $A\in\mathcal{N}_n$ 不可约且所有 $a_{ii}>0$, 则 $A^{n-1}>O$.

5. 设 $A\in\mathcal{N}_n(n\geqslant 2)$. 试证: A 不可约当且仅当 $\rho(A)$ 为其单重特征值且 A 与 A^{T} 分别有正特征向量对应于 $\rho(A)$.

6. 设 $A\in\mathcal{N}_n$. 试证: A 不可约当且仅当对每一对 (i,j), $1\leqslant i,j\leqslant n$, 存在正整数 $q\leqslant n$ (与 i,j 有关) 使得 $a_{ij}^{(q)}>0$, 这里, $a_{ij}^{(q)}$ 表示矩阵 A^q 的 (i,j) 元素.

7. 证明: 定理 13 中 $A\in\mathcal{N}_n(n\geqslant 2)$ 为不可约的当且仅当 (3) 成立.

8. 证明推论 18.

9. 试证: (18) 式中第一个等式可换为

$$\rho(A)=\max_{x\in H^n}\min_{x_i>0}\left(\frac{1}{x_i}\sum_{j=1}^n a_{ij}x_j\right).$$

但是该式中第二个等式不能换为

$$\rho(A)=\min_{x\in H^n}\max_{x_i>0}\left(\frac{1}{x_i}\sum_{j=1}^n a_{ij}x_j\right).$$

10. 设 $A\in\mathcal{N}_n$ 不可约. 试证:

$$\max_{x>0}\min_{y>0}\frac{\boldsymbol{y}^{\mathrm{T}}A\boldsymbol{x}}{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{x}}=\rho(A)=\min_{y>0}\max_{x>0}\frac{\boldsymbol{y}^{\mathrm{T}}A\boldsymbol{x}}{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{x}}.$$

11. 设 $A\in\mathcal{N}_n$ 且 $s>\rho(A)$. 试证: $(sI-A)^{-1}\geqslant O$; 若 A 还是不可约的, 则有 $(sI-A)^{-1}>O$.

12. 证明: 若 $b>c>0$, 且 $\boldsymbol{A}=(a_{ij})\in\boldsymbol{M}_n(\mathbb{C})$ 满足 $|a_{ij}|\leqslant b$, $\forall i<j$; $|a_{ij}|\leqslant c$, $\forall i>j$; $|a_{ii}|>(bc^{1/n}-cb^{1/n})/(b^{1/n}-c^{1/n})$, $\forall i$, 则 $\det A\neq 0$.

13. 设 $A=(a_{ij})$ 为 $n(n\geqslant 2)$ 阶正矩阵, 其 n 个行和分别为 $r_1,\cdots,r_n$. 令 $R=\max\limits_i r_i$, $r=$

$\min\limits_i r_i, \eta=\min\limits_{i,j}\alpha_{ij}$ 与 $\delta=(r-\eta)/(R-\eta)$. 证明：若 $R>r$，则有

$$r+\eta\left(\frac{1}{\sqrt{\delta}}-1\right)\leqslant\rho(A)\leqslant R-\eta(1-\sqrt{\delta}).$$

14. 设 $A=(a_{ij}), B=(b_{ij})\in\mathcal{N}_n$ 不可约，且 $B\geqslant A$. 令 $m=\min\limits_{i,j}(b_{ij}-a_{ij})$ 与 $M=\max\limits_{i,j}(b_{ij}-a_{ij})$. 试证：若 $R=(r_{ij})\in\mathcal{N}_n$ 与 A 或 B 可交换，且 R 所有列向量不为零向量，则有

$$\frac{m}{\max\limits_i \max\limits_j r_{ij}/\sum\limits_{k=1}^n r_{kj}}\leqslant\rho(B)-\rho(A)\leqslant\frac{M}{\min\limits_i \min\limits_j r_{ij}/\sum\limits_{k=1}^n r_{kj}}.$$

7.1.2 Perron-Frobenius 理论的进一步结果

由 7.1.1 小节定理 14 知道，不同于正方阵，一般非负不可约矩阵 A 可能有模为 $\rho(A)$ 的几个特征值. 本小节讨论非负不可约矩阵 Perron-Frobenius 理论的进一步结果，它们包括特征值结构以及其他更深入的结论.

定义 1 假如 $A\in\mathcal{N}_n$ 不可约，则模等于 $\rho(A)$ 的 A 特征值的个数 k 称为 A 的**循环指标**. 若 $k=1$，则 A 称为**素矩阵**(primitive matrix)；若 $k>1$，则 A 称为具有指标 $k(>1)$ 的**循环矩阵**(或**非素矩阵**).

按上述定义与 7.1.1 小节定理 14 中(1)，任意正方阵一定是素矩阵，但反之不真，如 $A=\begin{bmatrix}0 & 2\\1 & 1\end{bmatrix}$.

定理 2 设 $A\in\mathcal{N}_n$ 不可约，模为 $\rho(A)$ 的 A 的 k 个特征值为

$$\lambda_1=\rho(A)e^{i\theta_1}, \lambda_2=\rho(A)e^{i\theta_2}, \cdots, \lambda_k=\rho(A)e^{i\theta_k}, \tag{1}$$

这里，$0=\theta_1<\theta_2<\cdots<\theta_k<2\pi$. 则 $\lambda_1,\lambda_2,\cdots,\lambda_k$ 为方程 $\lambda^k-(\rho(A))^k=0$ 的 k 个不同根.

证明 应用 7.1.1 小节定理 7(取 $B=A$ 与 $\beta=\lambda_t$)，我们有

$$A=e^{i\theta_t}D_tAD_t^{-1}, \qquad |D_t|=I, t=1,\cdots,k, \tag{2}$$

于是，

$$A=e^{i(\theta_t\pm\theta_p)}(D_tD_p^{\pm1})A(D_tD_p^{\pm1})^{-1}, \qquad t,p=1,\cdots,k. \tag{3}$$

这表明 A 与 $e^{i(\theta_t\pm\theta_p)}A$ 相似，因而 $\rho(A)e^{i(\theta_t\pm\theta_p)}$ 是模为 $\rho(A)$ 的 A 的特征值，$t,p=1,\cdots,k$，这表明：$e^{i(\theta_t\pm\theta_p)}\in\{e^{i\theta_1},\cdots,e^{i\theta_k}\}$，$t,p=1,\cdots,k$. 因此，$e^{i\theta_1},\cdots,e^{i\theta_k}$ 构成一个有限乘法群(阶数为 k). 因有限群的每一个元素的阶都是群的阶的因子，所以对所有 $1\leqslant t\leqslant k$，$e^{ik\theta_t}=1$，于是，$\theta_t=2\pi t/k$，换言之，$\lambda_1,\cdots,\lambda_k$ 为方程 $\lambda^k-\rho(A)^k=0$ 的 k 个不同根. □

我们指出，满足(2)式两个条件的 D_t 不是唯一的. 但如约定 D_t 的头一个对角元素为 1，则 D_t 便为唯一的，因为若 $A\boldsymbol{x}=\rho(A)\boldsymbol{x}$ 与 $\boldsymbol{x}>\boldsymbol{0}$，令 $\boldsymbol{y}^{(t)}=D_t\boldsymbol{x}$，按(2)式则有 $A\boldsymbol{y}^{(t)}=\lambda_t\boldsymbol{y}^{(t)}$，$\boldsymbol{y}^{(t)}\neq\boldsymbol{0}$，由于 $\lambda_t\in\sigma(A)$ 为单重的，故 $\boldsymbol{y}^{(t)}$ 与 D_t 不计常数因子是唯一确定的. 这时，类似于前一定理的证明，我们可以看出：$D_1=I, D_2,\cdots,D_k$ 也构成

阶数为 k 的有限乘法群. 因此, $D_t^k=I, t=1,\cdots,k$. 在后面定理 5 证明中我们将应用这个事实.

定理 2 告诉我们,循环指标为 k 的非负不可约矩阵 A 有 k 个特征值等分复平面上以原点为圆心以 $\rho(A)$ 为半径的圆周,且其中有一点为 $(\rho(A),0)$. A 的其余 $n-k$ 个特征值都在这个圆的内部,它们的分布也呈明显的规律,此由下面定理给出.

定理 3 设 $A\in\mathcal{N}_n$ 不可约,其循环指标为 k. 则将复平面绕原点旋转角 $2\pi/k$,但不是更小的正角, A 的特征值集合保持不变.

证明 设 A 的特征值(计入重根)为 $\lambda_1,\lambda_2,\cdots,\lambda_n$. 则矩阵 $e^{i2\pi/k}A$ 的特征值应是 $e^{i2\pi/k}\lambda_1, e^{i2\pi/k}\lambda_2,\cdots,e^{i2\pi/k}\lambda_n$. 从定理 2 证明知道, A 与 $e^{i2\pi/k}A$ 相似,因而 A 的特征值集合应等于 $e^{i2\pi/k}A$ 的特征值集合. 由定义 1,模为 $\rho(A)$ 的 A 的特征值恰有 k 个,这一事实断定,对小于 $2\pi/k$ 的正角,旋转不变性不成立. □

前两个定理很好地刻画出循环指标为 k 的非负不可约矩阵 A 的谱结构:若 A 有特征值位于圆周 $|z|=r$ 上, $0<r\leqslant\rho(A)$,则此圆周上有 A 的 k 个不同的特征值,其重数相同,它们等分该圆周. 例如, A 为循环指标 $k=4$ 的六阶循环矩阵,则 A 的特征值为 $\pm\rho(A),\pm i\rho(A),0,0$. 一般地有

推论 4 设 $A\in\mathcal{N}_n$ 不可约,其循环指标为 k. 则

$$\det(\lambda I-A)=\lambda^m(\lambda^k-\rho(A)^k)(\lambda^k-\alpha_2\rho(A)^k)\cdots(\lambda^k-\alpha_r\rho(A)^k),$$

其中, $m\geqslant 0, r\geqslant 1, \alpha_1=1$,且 $m+rk=n$;当 $r>1$ 时对 $1<i\leqslant r$,都有 $|\alpha_i|<1$.

现在给出非负不可约矩阵的 Perron-Frobenius 理论的第二部分,其中(1)与(2)分别为前面证明过的定理 2 与定理 3.

定理 5 设 $A\in\mathcal{N}_n$ 不可约,其循环指标为 k.

(1) 设模为 $\rho(A)$ 的 A 的 k 个特征值是

$$\lambda_1=\rho(A)e^{i\theta_1},\lambda_2=\rho(A)e^{i\theta_2},\cdots,\lambda_k=\rho(A)e^{i\theta_k},$$

这里, $0=\theta_1<\theta_2<\cdots<\theta_k<2\pi$. 则 $\lambda_1,\lambda_2,\cdots,\lambda_k$ 为方程 $\lambda^k-(\rho(A))^k=0$ 的 k 个不同根.

(2) 当复平面绕原点旋转角 $2\pi/k$,但不是更小的正角时, A 的特征值集合保持不变.

(3) 若 $k>1$,则存在置换矩阵 P 使得 P^TAP 具有如下形式:

$$P^TAP=\begin{bmatrix} O & A_{12} & O & \cdots & O \\ O & O & A_{23} & \cdots & O \\ \vdots & \vdots & \vdots & & \vdots \\ O & O & O & \cdots & A_{k-1,k} \\ A_{k1} & O & O & \cdots & O \end{bmatrix}, \tag{4}$$

其中,块对角线上的零块都是方阵.

证明 只要证明(3). 令 $D=D_2$. 因 $D^k=I$(见定理 2 后面的说明),所以存在置

换矩阵 P 使得

$$P^{\mathrm{T}}DP = \begin{bmatrix} e^{i\delta_1}I_1 & O & \cdots & O \\ O & e^{i\delta_2}I_2 & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \cdots & e^{i\delta_s}I_s \end{bmatrix}, \tag{5}$$

其中，各 I_j 为单位矩阵但不必同阶；$\delta_j=(2\pi/k)n_j$，$0=n_1<n_2<\cdots<n_s<k$. 显然，(5)式中 $s>1$. 事实上，假若 $s=1$，则 $P^{\mathrm{T}}DP=I$，因而 $\boldsymbol{y}^{(2)}\equiv D_2\boldsymbol{x}=\boldsymbol{x}$（这里 $\boldsymbol{x}>\boldsymbol{0}$ 为 A 的 Perron 向量）. 这时，按(2)式，$A\boldsymbol{y}^{(2)}=AD_2\boldsymbol{x}=e^{i\theta_2}D_2A\boldsymbol{x}=\rho(A)e^{i\theta_2}\boldsymbol{x}$，这是不可能的，因为 $A\boldsymbol{y}^{(2)}=A\boldsymbol{x}$ 为实向量，而 $\rho(A)e^{i\theta_2}\boldsymbol{x}$ 不是实向量. 用同一置换矩阵 P 构成 $P^{\mathrm{T}}AP$，并按(5)式的分块法把 $P^{\mathrm{T}}AP$ 分为 s^2 个子矩阵 A_{pq}，即有

$$P^{\mathrm{T}}AP = [A_{pq}]_{p,q=1}^{s}. \tag{6}$$

从 $A=e^{i2\pi/k}DAD^{-1}$ 推出

$$P^{\mathrm{T}}AP = \varepsilon(P^{\mathrm{T}}DP)(P^{\mathrm{T}}AP)(P^{\mathrm{T}}DP)^{-1}, \qquad \varepsilon = e^{i2\pi/k}.$$

按(5)与(6)二式，并比较上式两端块元素得到 s^2 个方程式

$$A_{pq} = e^{i2\pi(1+n_p-n_q)/k}A_{pq}, \qquad 1\leqslant p,q\leqslant s.$$

上式推出，$A_{pp}=O$，$p=1,\cdots,s$，并且，若 $A_{pq}\neq O$，则有 $n_q\equiv 1+n_p \pmod{k}$. 另一方面，由于 A 不可约，故对每一个 $p(1\leqslant p\leqslant s)$，至少存在一个 $q=q(p)$，使得 $A_{pq}\neq O$. 现取 $p=1$，因 $n_1=0$，我们有 $n_{q(1)}\equiv 1 \pmod{k}$，于是，$q(1)$只能取一个值，又由于 $n_1<n_2<\cdots<n_s<k$ 都是整数，故有 $q(1)=2$ 与 $n_2=1$. 再取 $p=2$，因 $n_2=1$，我们有 $n_{q(2)}\equiv 2 \pmod{k}$，于是有 $q(2)=3$ 与 $n_3=2$. 用归纳法可以证明，当 $1\leqslant p\leqslant s-1$ 时，$n_p=p-1$ 与 $q(p)=p+1$. 因此，$A_{p,p+1}\neq O$ 且 $A_{pl}=O$，$l\neq p$，$p=1,\cdots,s-1$. 对于 $p=s$，对应的 $q(s)$满足 $n_{q(s)}\equiv 1+n_s \pmod{k}$，此推出 $1+n_s-n_{q(s)}=k$，因 $0\leqslant n_j<n_s<k$，$\forall 1\leqslant j\leqslant s-1$. 但 $n_1=0$，$n_2=1$，$\cdots$，$n_{s-1}=s-2$，故有 $n_s=k-1$ 与 $n_{q(s)}=0$，即 $q(s)=1$，于是 $A_{s1}\neq O$ 与 $A_{sl}=O$，$\forall l\neq 1$. 余下只要证：$s=k$. 取 $p=s-1$，则 A_{pl} 仅当 $l=s$ 时不为零. 因此，$1+n_{s-1}-n_s=0$. 但 $n_{s-1}=s-2$，且 $n_s=k-1$，于是，$s=k$. □

上述定理(3)给出循环指标大于 1 的非负不可约矩阵谱性质与矩阵零元素分布之间的一种关系. 通常将(4)式中分块矩阵称为具有循环指标 $k(>1)$的非负不可约 n 阶矩阵 A 的 **Frobenius 法式.**

假如在定理 5 证明中，取置换矩阵 P 使得(5)式中有 $0=n_s<\cdots<n_2<n_1<k$，则应用类似的推证可得

$$P^{\mathrm{T}}AP = \begin{bmatrix} O & O & \cdots & O & \widetilde{A}_{1k} \\ \widetilde{A}_{21} & O & \cdots & O & O \\ O & \widetilde{A}_{32} & \cdots & O & O \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ O & O & \cdots & \widetilde{A}_{k,k-1} & O \end{bmatrix}, \tag{7}$$

式中块对角线上的零块均为方阵.(7)式中矩阵有时也称为具有循环指标 $k(>1)$ 的非负不可约 n 阶矩阵 A 的 **Frobenius 法式**.

我们顺便地指出,具有形式(4)或(7)的非负矩阵可能是可约的. 例如,形式如(4)或(7)式的矩阵($k=2$)

$$A=\left[\begin{array}{cc:cc}0&0&0&1\\0&0&1&1\\ \hdashline 1&1&0&0\\1&0&0&0\end{array}\right]$$

是可约的. 另一方面,在分块矩阵(4)或(7)中,即使所有块为方阵且可约的,但矩阵 A 仍可能是不可约的. 例如形式如(4)或(7)式的矩阵($k=2$)

$$A=\left[\begin{array}{cc:cc}0&0&1&0\\0&0&1&1\\ \hdashline 1&1&0&0\\0&1&0&0\end{array}\right]$$

为非负不可约的,虽然它的所有块都是可约矩阵. 因此,(4)或(7)式不一定与非负或不可约矩阵联系在一起. 一般地,若 $A=(a_{ij})\in \boldsymbol{M}_n(\mathbb{C})(n\geqslant 2)$ 经同步变换 P 使得 $P^{\mathrm{T}}AP$ 有形式(7),其中块对角线上零块均为方阵,则称 A 为**指标 $k(>1)$ 的弱循环矩阵**. 但要注意,一个矩阵可以同时为不同指标的弱循环矩阵,因为存在不同的分块法. 例如对下面矩阵 A,按不同分块法,A 可以分别为指标 2,3,4 的弱循环阵:

$$A=\begin{bmatrix}0&0&0&-\mathrm{i}\\0&0&0&0\\0&1&0&0\\0&0&0&0\end{bmatrix}.$$

从前面讨论知道,非负不可约矩阵的循环指标 k 对了解该矩阵的谱结构至关紧要. 现在介绍确定它的一些方法(见定理 6 与定理 10)与其他有关结果.

定理 6 设 $A\in\mathscr{N}_n$ 不可约,其循环指标为 k,并且

$$\det(\lambda I-A)=\lambda^n+a_1\lambda^{n_1}+a_2\lambda^{n_2}+\cdots+a_h\lambda^{n_h},$$

式中,$a_1\neq 0,\cdots,a_h\neq 0,n>n_1>n_2>\cdots>n_h\geqslant 0$. 则

$$k=\mathrm{g.\,c.\,d.}\,(n-n_1,n_1-n_2,\cdots,n_{h-1}-n_h),\tag{8}$$

即 k 为 $n-n_1,n_1-n_2,\cdots,n_{h-1}-n_h$ 的最大公因数.

证明 设 $m\geqslant 1$ 为使得 A 与 $\mathrm{e}^{\mathrm{i}2\pi/m}A$ 相似的整数. 则 A 与 $\mathrm{e}^{\mathrm{i}2\pi/m}A$ 的特征多项式相等,即有

$$\lambda^n+a_1\lambda^{n_1}+\cdots+a_h\lambda^{n_h}=\lambda^n+a_1\theta^{n-n_1}\lambda^{n_1}+\cdots+a_h{}^{\theta^{n-n_h}}\lambda^{n_h},$$

其中 $\theta=\mathrm{e}^{\mathrm{i}2\pi/m}$. 因此,$a_t=a_t\theta^{n-n_t}(1\leqslant t\leqslant h)$,这表明 m 能整除 $n-n_1,n-n_2,\cdots,n-$

n_h. 反之，如 m 能整除 $n-n_1, n-n_2, \cdots, n-n_h$，将前面推证倒过来便得 A 与 θA 有相等的特征多项式，因而有相同的特征值. 但由定理 5 中(2)知道，A 与 $e^{i2\pi/m}A$ 当 $m=k$ 时有相同的特征值，而当 $m>k$ 时，二者没有相同的特征值集合. 连同前一部分证明我们得出

$$k = \mathrm{g.c.d}(n-n_1, n-n_2, \cdots, n-n_h),$$

由此容易得到公式(8). □

因为 $\mathrm{tr}A$ 为 $\det(\lambda I - A)$ 中 λ^{n-1} 系数的相反数，所以若此系数不为零，则定理 6 中 $n_1 = n-1$，因而 $k=1$. 这样一来，我们有如下判定非负不可约矩阵为素矩阵的简单办法.

推论 7 设 $A \in \mathcal{N}_n$ 不可约. 若 $\mathrm{tr}A > 0$，则 A 为素矩阵.

我们指出，这个结果也可从定理 5 中(3)直接推得，因为从(4)式看出，循环矩阵的迹必为零.

下面结果提供素矩阵的一个重要特征.

定理 8(Frobenius) 设 $A \in \mathcal{N}_n$. 则 A 为素矩阵当且仅当存在一个正整数 m 使得 $A^m > O$. 同时，若 $A^m > O$，则 $A^q > O$, $\forall q \geqslant m$.

证明 设 $A^m > O$ 对某正整数 m 成立. 则显然 $A \in \mathcal{N}_n$ 不可约. 假若 A 有 $k(>1)$ 个模为 $\rho(A)$ 的特征值，那么，A^m 也有 k 个模为 $\rho(A)^k$ 的特征值，这与 A^m 为正矩阵的事实抵触(见 7.1.1 小节定理 14 中(1)). 因此，A 为素矩阵. 反之，设 A 为素矩阵. 将谱分解定理用于 A^m，则有 $\lim\limits_{m\to\infty}(\rho(A)^{-1}A)^m = L > O$ (见习题 8)，因而对某个正整数 m 有 $(\rho(A)^{-1}A)^m > O$，此等价于 $A^m > O$. 最后一个结论显然成立，因为 $A^q = A^{q-m}A^m$ 为非负不可约矩阵(或单位矩阵)与正矩阵之积，所以 $A^q > O$, $\forall q \geqslant m$. □

由上述定理看出，若 $A \in \mathcal{N}_n$ 对某正整数 m 有 $A^m > O$，则对所有正整数 r，A^r 都是素矩阵. 反之，若对某个正整数 r，A^r 为素矩阵，则存在正整数 m 使得 $A^m > O$. 但是，对于具有指标 $k(>1)$ 的循环矩阵，它的乘幂可能是可约矩阵因而与素矩阵的乘幂之间有着本质上的差别. 确切地有

定理 9 设 $A \in \mathcal{N}_n$ 不可约，它有循环指数 $k>1$. 则存在 n 阶置换矩阵 P 使得对所有正整数 j，

$$P^{\mathrm{T}} A^{jk} P = \mathrm{diag}[C_1^j, C_2^j, \cdots, C_k^j], \tag{9}$$

其中每一个 C_i 为素矩阵，且有

$$\rho(C_1) = \rho(C_2) = \cdots = \rho(C_k) = (\rho(A))^k. \tag{10}$$

证明 设 $A \in \mathcal{N}_n$ 为具有指标 $k(>1)$ 的循环矩阵. 从(4)式经直接计算 $P^{\mathrm{T}}AP$ 的乘幂看出，(9)式对所有正整数 j 成立，且其中的矩阵 $C_i = A_{i,i+1}A_{i+1,i+2}\cdots A_{i-1,i}$ (k 个矩阵的乘积)，$i=1,\cdots,k$(这里约定：$A_{01} = A_{k1}$, $A_{k,k+1} = A_{k1}$，等等). 容易验证各 C_i 为非负的方阵，其阶数等于(4)式中矩阵块对角线上第 i 个零块的阶数. 并且

按第一章 1.1.1 小节特征值性质(4)，各 C_i 有相同的谱半径 $\rho(C_i)$. 余下只要证明：各 C_i 为素矩阵. 考虑 A 的 Frobenius 法式(4). 设模为 $\rho(A)$ 的 A 的特征值为

$$\lambda_t = \rho(A)e^{i2\pi t/k}, \qquad t = 0,1,\cdots,k-1, \tag{11}$$

且 $\boldsymbol{x}=(x_1,\cdots,x_n)^{\mathrm{T}}>\boldsymbol{0}$ 为 A 的 Perron 向量. 则 A^k(因而 $P^{\mathrm{T}}A^kP$)的模为 $(\rho(A))^k$ 的特征值有 k 个(计入重数)，且按(11)式它们都等于 $(\rho(A))^k$. 因此，$\rho(C_1)=\cdots=\rho(C_k)=(\rho(A))^k$，且每一个 $\rho(C_j)(1\leqslant j\leqslant k)$ 为 C_j 的模为 $\rho(C_j)$ 的唯一单重特征值. 现将正向量 $P^{\mathrm{T}}\boldsymbol{x}$ 写为 $P^{\mathrm{T}}\boldsymbol{x}=\sum_{j=1}^{k}{}^{\boldsymbol{\cdot}}\ \boldsymbol{x}^j$，使得此分块列向量与 A 的 Frobenius 法式的分块匹配. 于是，$A\boldsymbol{x}=\rho(A)\boldsymbol{x}$ 蕴涵 $(P^{\mathrm{T}}AP)(P^{\mathrm{T}}\boldsymbol{x})=\rho(A)(P^{\mathrm{T}}\boldsymbol{x})$ 与 $(P^{\mathrm{T}}A^kP)\cdot(P^{\mathrm{T}}\boldsymbol{x})=\rho(A)^kP^{\mathrm{T}}\boldsymbol{x}$，进而

$$(P^{\mathrm{T}}A^kP)(P^{\mathrm{T}}\boldsymbol{x}) = \sum_{j=1}^{k}{}^{\boldsymbol{\cdot}}\ C_j\boldsymbol{x}^j = \sum_{j=1}^{k}{}^{\boldsymbol{\cdot}}\ (\rho(A))^k\boldsymbol{x}^j.$$

因此，

$$C_j\boldsymbol{x}^j = (\rho(A))^k\boldsymbol{x}^j, j = 1,\cdots,k.$$

这表明对应于 $\rho(C_j^{\mathrm{T}})=(\rho(A))^k$，$C_j$ 有正特征向量 $\boldsymbol{x}^j$，$j=1,\cdots,k$. 类似地可以证明：各个 C_j^{T} 有正特征向量对应于 $\rho(C_j^{\mathrm{T}})=(\rho(A))^k$. 应用习题 7.1.1 题 5 便知各 C_j 为(非负)不可约的. 连同前面证明的结果可以断定各 C_j 为素矩阵. □

现在还回到确定非负不可约矩阵的循环指标 k 的求值问题. 除了定理 6 以外，非负不可约矩阵的循环指标 k 也可以由矩阵有向图确定，此种方法既直观又简便. 考虑 $A=(a_{ij})\in\boldsymbol{M}_n(\mathbb{C})$ 的有向图 $\Gamma(A)$. 若 $\boldsymbol{v}=\{\widehat{i_0i_1},\widehat{i_1i_2},\cdots,\widehat{i_{l-1}i_l}\}$ 为 $\Gamma(A)$ 的一条由 l 个有向弧组成的有向路，则称 l 为 $\boldsymbol{v}$ 的**长度**. 当 $i_l=i_0$ 时，则称 $\boldsymbol{v}$ 为 $\Gamma(A)$ 的一条**回路**.(与第六章 6.1.2 小节中简单回路不同，除了 $i_l=i_0$ 以外这里允许 $\boldsymbol{v}$ 有相重的顶点)对于 $A=(a_{ij})\in\mathcal{N}_n$，其有向图 $\Gamma(A)$ 与 $A^r=(a_{ij}^{(r)})\in\mathcal{N}_n$ 有向图 $\Gamma(A^r)$ 之间有着简单的联系，其中，r 为正整数. 若 $\Gamma(A)$ 中有一条长度为 r 的有向路自顶点 i 到顶点 j，即有 $a_{il_1}a_{l_1l_2}\cdots a_{l_{r-1}j}>0$，那么，$\Gamma(A^r)$ 有一条有向弧自顶点 i 到 j.(这是因为 $A\geqslant O$，且 $a_{ij}^{(r)}=a_{il_1}a_{l_1l_2}\cdots a_{l_{r-1}j}+$(非负数)，这里，(非负数)表示若干个乘积 $a_{im_1}a_{m_1m_2}\cdots a_{m_{r-1}j}$ 之和，所以 $a_{ij}^{(r)}>0$)换句话说，$\Gamma(A^r)(r\geqslant1)$ 每一个连接顶点 i 与 j 的有向弧相当于 $\Gamma(A)$ 中连接顶点 i 与 j 的一条长度为 r 的有向路.

定理 10(Romanovsky)　设 $A\in\mathcal{N}_n$ 不可约，它的有向图为 $\Gamma(A)$. 若 $\mathcal{S}_i$ 表示 $\Gamma(A)$ 中通过顶点 i 的所有回路的长度 $\boldsymbol{v}_i$ 的集合，且

$$k_i = \underset{\boldsymbol{v}_i\in\mathcal{S}_i}{\mathrm{g.c.d.}}\{\boldsymbol{v}_i\}, \qquad i = 1,\cdots,n, \tag{12}$$

则 $k_1=k_2=\cdots=k_n=k$，其中 k 为 A 的循环指标.

证明　首先证明：$\mathcal{S}_i$ 为加法半群，$i=1,\cdots,n$. 设 $A^r=(a_{ij}^{(r)})(r\geqslant1)$，且 $\boldsymbol{v}'_i$，$\boldsymbol{v}''_i\in\mathcal{S}_i$. 按前面介绍的有向图知识，$a_{ii}^{(v'_i)}$ 与 $a_{ii}^{(v''_i)}$ 均为正数. 因此，从

$$a_{ii}^{(v'_i+v''_i)} = \sum_{l=1}^{n} a_{il}^{(v'_i)} a_{li}^{(v''_i)} \geqslant a_{ii}^{(v'_i)} a_{ii}^{(v''_i)} > 0$$

便得 $\boldsymbol{v}'_i+\boldsymbol{v}''_i\in\mathscr{S}_i, i=1,\cdots,n$. 这样一来，除了最大公因数 k_i 的有限个倍数以外，$\mathscr{S}_i$ 含有 k_i 的所有倍数. 假如某 $k_i=1$，则对所有充分大的 p 有 $a_{ii}^{(p)}>0$，因而 A 为素矩阵，因为如 A 为循环矩阵，由它的 Frobenius 法式看出，$a_{ii}^{(p)}>0$ 不能对所有充分大的 p 都成立. 因此按定理 8，对所有充分大的 p 有 $a_{jj}^{(p)}>0, j=1,\cdots,n$，因而按 k_j 的定义，$k_1=k_2=\cdots=k_n=1$. 现在假定某个 $k_i>1$. 这意味着对所有充分大的 p，$a_{ii}^{(p)}$ 不全为正数（否则 $k_i=1$），因而 A 必定为循环矩阵. 令它的循环指标为 $k(>1)$. 从定理 9 的证明看出，有正对角元素的 A 的乘幂仅是 A^{jk}（j 为正整数），且(9)式中所有 $C_i(1\leqslant i\leqslant k)$ 均为素矩阵. 于是，对所有充分大的 j，A^{jk} 的所有对角元素为正数（见定理 8）. 按 k_i 的定义，这个事实蕴涵 $k_1=k_2=\cdots=k_n=k$. □

为了例证上述定理的结论，考虑下列非负矩阵：

$$A_1=\begin{bmatrix}0&1&0&0\\0&0&1&0\\0&0&0&1\\1&0&0&0\end{bmatrix},\qquad A_2=\begin{bmatrix}0&1&0&0\\0&0&1&0\\0&0&0&1\\1&1&0&0\end{bmatrix},$$

$$A_3=\begin{bmatrix}0&0&1&0\\0&0&0&1\\1&0&0&0\\0&1&0&0\end{bmatrix},\qquad A_4=\begin{bmatrix}1&1&0&0\\0&0&1&0\\0&0&0&1\\1&0&0&0\end{bmatrix}.$$

它们的有向图分别如图 7.1 所示. 这里除了 $\Gamma(A_3)$ 以外，其余三个有向图均为强连接的，因而 $A_3\in\mathscr{N}_4$ 可约，但 $A_1,A_2,A_4\in\mathscr{N}_4$ 不可约. 对于 A_1 应用定理 10 有

$$k_1 = \text{g. c. d.}\{4,8,12,\cdots\} = 4,$$

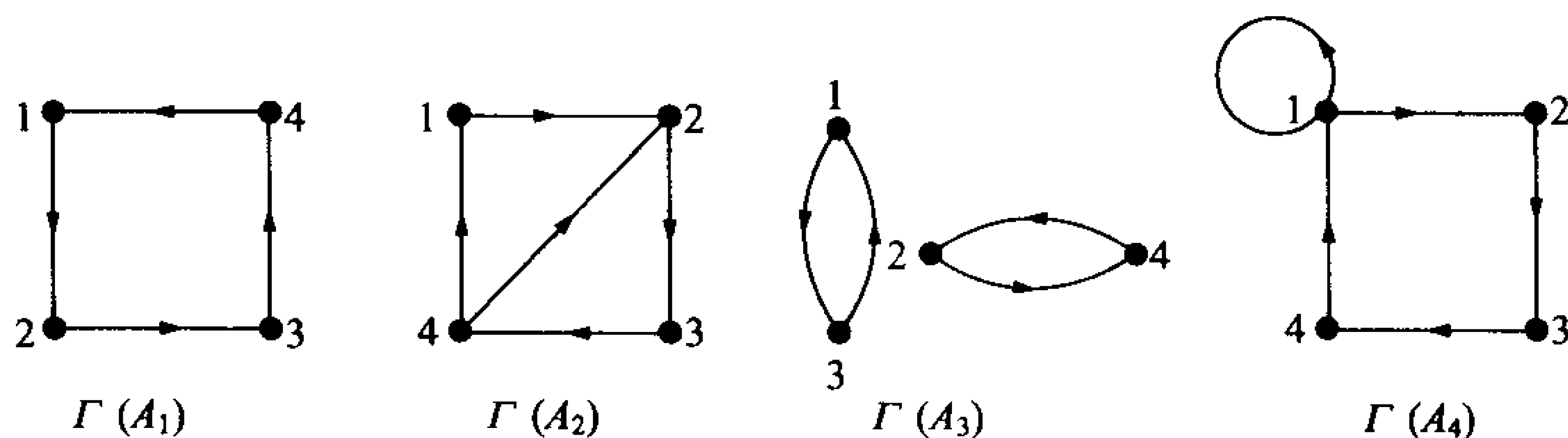


图 7.1　有向图

即 A_1 为指标 $k=4$ 的循环矩阵. 对于 A_2 有

$$k_4 = \text{g. c. d.}\{3,4,7,\cdots\} = 1,$$

即 A_2 为素矩阵. 对于 A_4，它有一个过顶点 1 的自身回路（长度为 1），故 $k_1=1$，即 A_4 也是素矩阵（亦可应用推论 7）.

最后，我们简单地介绍一下素矩阵的**素性指标**的概念.

定义 11　设 $A\in\mathcal{N}_n$ 为素矩阵. 则称使得 $A^m>O$ 的最小正整数 m 为 A 的**素性指标**(index of primitivity)，记为 $\gamma(A)$.

从习题 7.1.1 题 4 看出，若 $A\in\mathcal{N}_n(n\geqslant2)$ 为素矩阵且所有 $a_{ii}>0$，则有 $\gamma(A)\leqslant n-1$. 一般地，$\gamma(A)\leqslant n^n(n-1)$，这由下面定理保证.

定理 12　设 $A\in\mathcal{N}_n(n\geqslant2)$. 若 A 为素矩阵，则 $\gamma(A)\leqslant n^n(n-1)$.

证明　由于 $A\in\mathcal{N}_n$ 不可约，且对所有正整数 r，A^r 也是素矩阵，故按习题 7.1.1 题 6，存在正整数 r_1 使得 $a_{11}^{(r_1)}>0$. 因 $A^{r_1}\in\mathcal{N}_n$ 不可约，同理有正整数 r_2 使得 $a_{22}^{(r_1r_2)}>0$，这里，$a_{ij}^{(r_1r_2)}$ 表示 $(A^{r_1})^{r_2}$ 的 (i,j) 元素. 此时，$a_{11}^{(r_1r_2)}\geqslant(a_{11}^{r_1})^{r_2}>0$. 按这样方法继续进行下去，可以知道存在正整数 $r_1,r_2,\cdots,r_n$ 使得 $A^{r_1r_2\cdots r_n}\geqslant O$ 为素矩阵，且它的所有对角元素为正数. 再引用习题 7.1.1 题 4 便得，$A^m>O$，其中 $m=r_1r_2,\cdots,r_n(n-1)$. 但由习题 7.1.1 题 6 知道，$r_j\leqslant n(1\leqslant j\leqslant n)$，因而 $\gamma(A)\leqslant n^n(n-1)$. □

我们指出，上述定理给出的 $\gamma(A)$ 的上界可进一步地改进为 n^2-2n+2(见习题 10).

习题 7.1.2

1. 试问：

(1) 两个同阶素矩阵之积为素矩阵吗？

(2) 两个同阶素矩阵之积是非负不可约矩阵吗？

(3) 两个非负可约矩阵的乘积一定是非负可约矩阵吗？

2. 设 $A\in\mathcal{N}_n$ 为素矩阵，且 $B\in\mathcal{N}_n$. 证明：$A+B$ 为素矩阵.

3. 设 $A\in\mathcal{N}_n$ 为具有指标 $k>1$ 的循环矩阵，且它的 Frobenius 法式为(4)式. 试用这个法式直接证明定理 3 中循环矩阵 A 的特征值集合旋转不变性结果.

4. 设 $A-(a_{ij})\in\boldsymbol{M}_n(\mathbb{C})$ 为指标 $k>1$ 的弱循环矩阵. 试证：A 的特征多项式有形式

$$\det(tI-A)=t^m\prod_{i=1}^{r}(t^k-\sigma_i^k),\qquad m\geqslant0.$$

5. 设 A 同题 4. 试证：存在 n 阶置换矩阵 P 使得(9)式成立，其中每一个 C_i 为方阵，满足(10)式.

6. 应用本小节中介绍的有向图知识，试重新证明习题 7.1.1 题 6.

7. 设

$$B=\begin{bmatrix}0&1&&&&\\1&0&1&&&0\\&1&0&1&&\\&&\ddots&\ddots&\ddots&\\&0&&&&1\\&&&&1&0\end{bmatrix}\in\boldsymbol{M}_n(\mathbb{R}),\qquad n\geqslant2.$$

试证：$B\in\mathcal{N}_n$ 不可约，其循环指标为 2. 更一般地，设

$$B=\begin{bmatrix} O & B_{12} & & & \\ B_{21} & O & B_{23} & & O \\ & \ddots & \ddots & \ddots & \\ O & & & & B_{N-1,N} \\ & & & B_{N,N-1} & O \end{bmatrix},$$

其中，$N\geqslant 2$，且块对角线上零块皆为方阵. 试证：B 为指标 2 的弱循环阵.

8. 试证：若 $A\in\mathcal{N}_n$ 为素矩阵，则 $\lim\limits_{m\to\infty}(\rho(A)^{-1}A)^m=L>O$，这里，$L=\boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}$，$A\boldsymbol{x}=\rho(A)\boldsymbol{x}$，$\boldsymbol{x}>\boldsymbol{0}$，$A^{\mathrm{T}}\boldsymbol{y}=\rho(A)\boldsymbol{y}$，$\boldsymbol{y}>\boldsymbol{0}$ 且 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}=1$. 并且，若 $\lambda_{n-1}\in\sigma(A)$ 满足 $|\lambda|\leqslant|\lambda_{n-1}|$，$\forall\lambda\in\sigma(A)$ 但 $\lambda\neq\rho(A)$，$|\lambda_{n-1}|/\rho(A)<r<1$，则存在常数 c（与 r,A 有关）使得 $\|(\rho(A)^{-1}A)^m-L\|_\infty\leqslant cr^m$，$m=1,2,\cdots$.

9. 设 $A\in\mathcal{N}_n$ 为素矩阵. 试证：$\lim\limits_{m\to\infty}[a_{ij}^{(m)}]^{1/m}=\rho(A)$，$\forall i,j\in\mathcal{N}$，这里，$a_{ij}^{(m)}$ 为 A^m 的 (i,j) 元素，并且，

$$\lim_{m\to\infty}(\mathrm{tr}A^m)^{1/m}=\rho(A).$$

10. 试证：若 $A\in\mathcal{N}_n(n\geqslant 2)$，则 A 为素矩阵当且仅当 $A^{n^2-2n+2}>O$.

11. 应用有向图证明 **Wielandt 矩阵**

$$A=\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & & 0 & 1 \\ 1 & 1 & 0 & \cdots & 0 & 0 \end{bmatrix}\in\mathcal{N}_n,\qquad n\geqslant 3$$

为不可约素矩阵. 此外，证明它的素性指标 $\gamma(A)=n^2-2n+2$.（因此，题 10 中，n^2-2n+2 为可达到的指数）.

*12. 设 $A\in\mathcal{N}_n$ 为素矩阵，且 A 的有向图 $\Gamma(A)$ 中最短简单回路有长度 s. 试证：A 的素性指标 $\gamma(A)\leqslant n+s(n-2)$.

13. 设 $A\in\mathcal{N}_n$ 不可约，$A\boldsymbol{x}=\rho(A)\boldsymbol{x}$，$\boldsymbol{x}>\boldsymbol{0}$，$A^{\mathrm{T}}\boldsymbol{y}=\rho(A)\boldsymbol{y}$，$\boldsymbol{y}>\boldsymbol{0}$，$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}=1$ 与 $L=\boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}$. 试证：

$$\lim_{N\to\infty}\frac{1}{N}\sum_{m=1}^{N}(\rho(A)^{-1}A)^m=L.$$

并且，有正常数 $c=c(A)$ 使得

$$\left\|\frac{1}{N}\sum_{m=1}^{N}(\rho(A)^{-1}A)^m-L\right\|_\infty\leqslant c/N,\qquad N=1,2,\cdots.$$

试对 $A=\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}$ 计算上述极限矩阵 L.

7.2 一般非负矩阵的情形

7.2.1 一般非负矩阵 Perron-Frobenius 理论的古典结果

7.1 节中介绍的非负不可约矩阵的某些结果可以用连续性推证方法扩充到一

般非负方阵的情形. 一个基本结论是,若 $A\in\mathcal{N}_n$ 则 $\rho(A)\in\sigma(A)$,且有 $\boldsymbol{x}\in\boldsymbol{E}^n$ 使得 $A\boldsymbol{x}=\rho(A)\boldsymbol{x}$. 事实上,每一个 $A\in\mathcal{N}_n$ 可以表示为 $A=\lim\limits_{m\to\infty}A_m$,这里,$A_m$ 为非负不可约的,甚至可选择为正矩阵,譬如,可取 $A_m=A+\frac{1}{m}J$,这里 J 表示所有元素皆为 1 的 n 阶矩阵,$m=1,2,\cdots$. 显然,$O<A_{m+1}<A_m$, $\forall m\geqslant 1$,因而 7.1.1 小节推论 9 断言,$0<\rho(A_{m+1})<\rho(A_m)$, $\forall m\geqslant 1$,于是 $\rho(A)=\lim\limits_{m\to\infty}\rho(A_m)\geqslant 0$. 现令 $\boldsymbol{x}^m\in\boldsymbol{E}^n$ 为正矩阵 A_m 的 Perron 向量,$m=1,2,\cdots$, 且令 $\boldsymbol{x}$ 为 $\{\boldsymbol{x}^m\}$ 某一收敛子列的极限. 则有 $A\boldsymbol{x}=\rho(A)\boldsymbol{x}$,$\boldsymbol{x}\in\boldsymbol{E}^n$,即 $\boldsymbol{x}$ 为对应于 $\rho(A)$ 的 A 的特征向量. 诸如上述的极限过程通常称为**连续性推证**. 但为进一步讨论一般非负方阵的 Perron-Frobenius 理论,还需要考察一下一般可约矩阵的构造. 设 $A=(a_{ij})\in\boldsymbol{M}_n(\mathbb{C})$ 可约,$n\geqslant 2$. 由定义知道,存在置换矩阵 P_1 使得

$$P_1AP_1^{\mathrm{T}}=\begin{bmatrix}B & C\\ O & D\end{bmatrix}\tag{1}$$

其中,B 与 D 为方阵. 若 B 或 D 仍为可约的,可以像在前一步骤里约化 A 那样去约化 B 或 D. 如此进行下去,有限步以后便知存在 n 阶置换矩阵 P,它是有限多个置换矩阵之积,使得

$$PAP^{\mathrm{T}}=\begin{bmatrix}A_{11} & A_{12} & \cdots & A_{1m}\\ O & A_{22} & \cdots & A_{2m}\\ \vdots & & \ddots & \vdots\\ O & O & \cdots & A_{mm}\end{bmatrix}\quad (m\geqslant 2),\tag{2}$$

其中块对角线上每块 A_{jj} $(1\leqslant j\leqslant m)$或为不可约矩阵,或为一阶零矩阵. (2)式称为**可约矩阵 $\boldsymbol{A}$ 的法式**. 当 $A=(a_{ij})\in\boldsymbol{M}_n(\mathbb{C})$不可约或为一阶零矩阵时,为统一起见,记 $A=[A_{11}]$.

可约矩阵 $A\in\boldsymbol{M}_n(\mathbb{C})$ $(n\geqslant 2)$的法式一般不是唯一的,但其对角块 A_{jj}(若不计同步变换)是唯一的,这是因为每一个不可约的 A_{jj} 与 A 的有向图 $\Gamma(A)$的强连接有向子图之间存在着一一对应. 例如五阶非负可约矩阵

$$A=\begin{bmatrix}1 & 6 & 0 & 0 & 1\\ 0 & 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 2 & 0\\ 0 & 4 & 0 & 0 & 0\\ 1 & 0 & 0 & 0 & 0\end{bmatrix}\tag{3}$$

有法式

$$PAP^{\mathrm{T}}=\left[\begin{array}{cc:ccc}1 & 1 & 6 & 0 & 0\\ 1 & 0 & 0 & 0 & 0\\ \hdashline 0 & 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 0 & 2\\ 0 & 0 & 4 & 0 & 0\end{array}\right]$$

即有

$$A_{11}=\begin{bmatrix}1 & 1\\ 1 & 0\end{bmatrix},\quad A_{12}=\begin{bmatrix}6 & 0 & 0\\ 0 & 0 & 0\end{bmatrix},\quad A_{22}=\begin{bmatrix}0 & 1 & 0\\ 0 & 0 & 2\\ 4 & 0 & 0\end{bmatrix}.$$

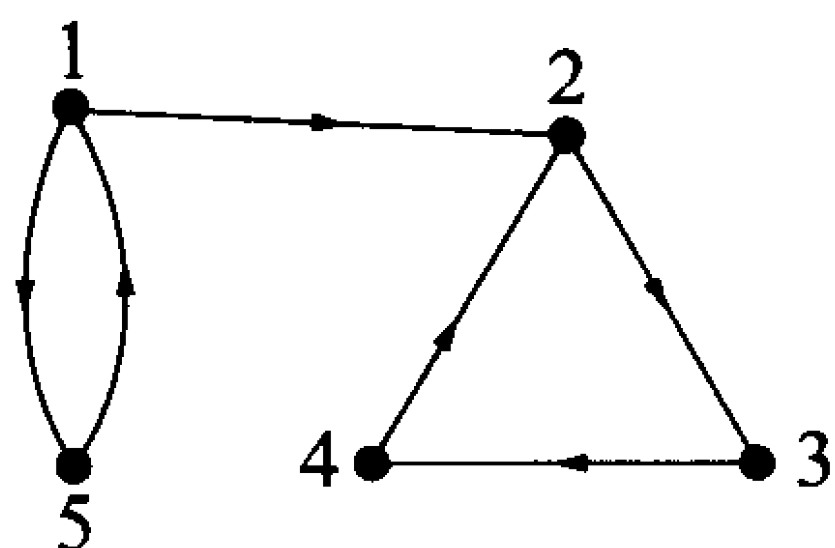


图 7.2　有向图 $\Gamma(A)$

这里，A_{11} 与 A_{22} 分别对应于 A 的有向图 $\Gamma(A)$ 的两个强连接有向子图(图 7.2).

现设 $A\in\mathcal{N}_n$ 有法式(2). 显然，A 的特征值集合由非负不可约矩阵或一阶零矩阵 A_{jj} $(1\leqslant j\leqslant m)$ 的所有特征值组成. 因此，我们有如下定理.

定理 1　设 $A\in\mathcal{N}_n$ 有法式(2). 则

(1) $\rho(A)\geqslant 0$ 为 A 的特征值，且除了 A 可约法式(2)为严格上三角形矩阵情形以外，$\rho(A)>0$.

(2) 对应于 $\rho(A)$，A 与 A^{T} 有非负的特征向量. A 的其他特征值一定没有正的特征向量.

证明　若 $\rho(A)=0$，则 $A\in\mathcal{N}_n$ 可约且(2)式中每个 A_{jj} $(1\leqslant j\leqslant m)$ 为一阶零矩阵，因而(2)式中矩阵为严格上三角形矩阵. (2)中第一部分前面已证. 现设 $\rho(A)\neq\lambda\in\sigma(A)$，且有 $\boldsymbol{v}\neq\boldsymbol{0}$ 满足 $A\boldsymbol{v}=\lambda\boldsymbol{v}$. 我们要证：$\boldsymbol{v}$ 不是正向量. 事实上，设 $\boldsymbol{u}\geqslant 0$ 为对应于 $\rho(A)$ 的 A^{T} 的特征向量，则

$$\lambda(\boldsymbol{v},\boldsymbol{u})=(A\boldsymbol{v},\boldsymbol{u})=(\boldsymbol{v},A^{\mathrm{T}}\boldsymbol{u})=\rho(A)(\boldsymbol{v},\boldsymbol{u}).$$

但 $\lambda\neq\rho(A)$，故有 $(\boldsymbol{v},\boldsymbol{u})=0$，这表明 $\boldsymbol{v}$ 不能是正向量. □

对应于 7.1.1 小节定理 7 中(1)，有如下的推广结果.

定理 2　设 $A\in\mathcal{N}_n$ 且 $|B|\leqslant A$. 则 $\rho(B)\leqslant\rho(A)$.

证明　只要考虑 $A\in\mathcal{N}_n$ 可约且有法式(2)的情形. 这时 $O\leqslant|PBP^{\mathrm{T}}|\leqslant P|B|P^{\mathrm{T}}\leqslant PAP^{\mathrm{T}}$. 将 PBP^{T} 写成分块形式 $[B_{ij}]_{i,j=1}^{m}$，使得 B_{jj} 与(2)式中 A_{jj} 为同阶矩阵，$j=1,\cdots,m$. 于是，$|B_{jj}|\leqslant A_{jj}$ $(j=1,\cdots,m)$，且 PBP^{T} 也是分块上三角形矩阵. 因此，按 7.1.1 小节定理 7，$\rho(B_{jj})\leqslant\rho(A_{jj})$ $(j=1,\cdots,m)$，因而 $\rho(B)=\max\limits_{j}\rho(B_{jj})\leqslant\max\limits_{j}\rho(A_{jj})=\rho(A)$. □

推论 3　若 B 为 $A\in\mathcal{N}_n$ 的主子阵，则 $\rho(B)\leqslant\rho(A)$. 并且，当 A 的任一元素增大时，$\rho(A)$ 不减小.

推论 4　设 $A\in\mathcal{N}_n$，$n\geqslant 2$. 则 A 可约当且仅当 $\rho(A)$ 为 A 的某个真正主子阵的

特征值.

证明　设 A 可约. 则有置换矩阵 P_1 使得(1)式成立，因而 $\rho(A)$ 或为 B 或为 D 的特征值. 逆命题成立由 7.1.1 小节推论 8 保证，因为当 $A\in\mathcal{N}_n(n\geqslant 2)$ 不可约时，$\rho(A)$ 一定大于 A 的任意真正主子阵的特征值. □

下列结论为非负不可约矩阵谱半径性质(见 7.1.1 小节推论 18)的部分推广.

定理 5　设 $A\in\mathcal{N}_n$. 则当且仅当 $\rho(A)\geqslant\alpha$ 时，存在某 n 维非零列向量 $\boldsymbol{x}\geqslant\boldsymbol{0}$ 使得 $A\boldsymbol{x}\geqslant\alpha\boldsymbol{x}$.

证明　设 $\rho(A)\geqslant\alpha$，按定理 1 中(2)，有 $\boldsymbol{x}\geqslant\boldsymbol{0}(\neq\boldsymbol{0})$ 使得 $A\boldsymbol{x}=\rho(A)\boldsymbol{x}\geqslant\alpha\boldsymbol{x}$. 反之，设 $A\boldsymbol{x}\geqslant\alpha\boldsymbol{x}$，$\boldsymbol{x}\geqslant\boldsymbol{0}(\neq\boldsymbol{0})$，$\alpha\in\mathbb{R}$. 取 $A_m=A+\frac{1}{m}J$，按本小节开始介绍的连续性推证方法与习题 7.1.1 题 9 容易推出，$\rho(A_m)\geqslant\alpha$. 令 $m\to\infty$，便有 $\rho(A)\geqslant\alpha$. □

对于非负矩阵谱半径的估计，应用连续性推证与前一定理结果，我们可以得到 7.1.1 小节定理 17 的推广.

定理 6　设 $A\in\mathcal{N}_n$ 且 $\boldsymbol{x}=(x_1,\cdots,x_n)^{\mathrm{T}}>\boldsymbol{0}$. 则有

$$\min_{1\leqslant i\leqslant n}\left(\frac{1}{x_i}\sum_{j=1}^{n}a_{ij}x_j\right)\leqslant\rho(A)\leqslant\max_{1\leqslant i\leqslant n}\left(\frac{1}{x_i}\sum_{j=1}^{n}a_{ij}x_j\right),\tag{4}$$

$$\rho(A)=\max_{y\in H^n}\min_{y_i>0}\left(\frac{1}{y_i}\sum_{j=1}^{n}a_{ij}y_j\right).\tag{5}$$

本定理证明留给读者(见习题 4).

由 7.1.1 小节推论 11 知道，若 $A\in\mathcal{N}_n$ 不可约与 $r=\rho(A)$，则 $B(\lambda)=\mathrm{adj}(\lambda I-A)=(b_{ij}(\lambda))$ 满足 $B(r)>O$. 对于一般非负可约矩阵情形，我们有

定理 7　设 $A\in\mathcal{N}_n$，$r=\rho(A)$. 则有 $B(\lambda)\geqslant O$ 与 $\dfrac{\mathrm{d}B(\lambda)}{\mathrm{d}\lambda}\geqslant O$，$\forall\lambda\geqslant r$.

证明　应用 7.1.1 小节推论 11 与连续性推证方法可得 $B(r)\geqslant O$. 对 $n=1$，$B(\lambda)\equiv I$ 与 $\mathrm{d}B(\lambda)/\mathrm{d}\lambda\equiv O$. 现假定本定理结果对阶数小于 n 的非负矩阵皆成立，考虑 n 阶非负矩阵 A 的情形. 令 $B^{(k)}(\lambda)\equiv\mathrm{adj}[(\lambda I-A)(k|k)]=(b_{ij}^{(k)}(\lambda))$ $(i,j=1,\cdots,k-1,k+1,\cdots,n)$. 将行列式 $\Delta(\lambda)=\det(\lambda I-A)$ 按第 k 行与第 k 列展开得出

$$\Delta(\lambda)=(\lambda-a_{kk})b_{kk}(\lambda)-\sum_{\substack{i,j=1\\ i,j\neq k}}^{n}b_{ji}^{(k)}(\lambda)a_{ik}a_{kj}.\tag{6}$$

用 r_k 表示多项式 $b_{kk}(\lambda)$ 的最大非负根(即 $r_k=\rho(A(k\mid k))$，$\forall k$). 按归纳假定，$B^{(k)}(r_k)\geqslant O$. 取 $\lambda=r_k$，则(6)式蕴涵 $\Delta(r_k)\leqslant 0$，于是，$r_k\leqslant r=\rho(A)$，$k=1,\cdots,n$. 同时，$\mathrm{d}b_{ij}(\lambda)/\mathrm{d}\lambda=\sum\limits_{\substack{k=1\\ k\neq i,j}}^{n}b_{ij}^{(k)}(\lambda)$. 按归纳假定，$B^{(k)}(\lambda)\geqslant O$，$\forall\lambda\geqslant r_k$. 因此，对 $\lambda\geqslant r$，$\mathrm{d}B(\lambda)/\mathrm{d}\lambda=(\mathrm{d}b_{ij}(\lambda)/\mathrm{d}\lambda)\geqslant O$，因而由 $B(r)\geqslant O$ 推得 $B(\lambda)\geqslant O$，$\forall\lambda\geqslant r$.

□

习题 7.2.1

1. 设

$$A=\begin{bmatrix}0&0&1&0\\1&1&1&3\\1&0&2&0\\2&2&0&1\end{bmatrix}\in\mathcal{N}_4.$$

写出 A 的可约法式，并求出 $\rho(A)$.

2. 证明：若 $A\in\mathcal{N}_n$，则当且仅当 A 的有向图 $\boldsymbol{\Gamma}(A)$ 至少有一回路时，$\rho(A)>0$.

3. 证明：若 $A=(a_{ij})\in\boldsymbol{M}_n(\mathbb{R})$，且 $a_{ij}>0$，$\forall\, i\neq j$，则 A 的有最大实部的特征值 ρ 为实数且重数等于 1，并且有 A 的正特征向量对应于 ρ. 若仅假定 $a_{ij}\geqslant 0$，$\forall\, i\neq j$，其对应的结论是什么？

4. 证明定理 6.

5. 设 $A\in\mathcal{N}_n$. 试证：$\rho((A+A^{\mathrm{T}})/2)\geqslant\rho(A)$，并且等式成立当且仅当 A 与 A^{T} 有公共的特征向量对应于 $\rho(A)$.

6. 设 $A\in\mathcal{N}_n$ 与 $\boldsymbol{x}>\boldsymbol{0}$. 试证：若 $\alpha\boldsymbol{x}\leqslant A\boldsymbol{x}\leqslant\beta\boldsymbol{x}$，$\alpha,\beta\in\mathbb{R}$，则 $\alpha\leqslant\rho(A)\leqslant\beta$；若 $\alpha\boldsymbol{x}<A\boldsymbol{x}\,(A\boldsymbol{x}<\beta\boldsymbol{x})$，则 $\alpha<\rho(A)\,(\rho(A)<\beta)$.

7. 设 $A\in\mathcal{N}_n$ 且 $\boldsymbol{x}^{(0)}\in\mathbb{R}^n$ 为任意正向量. 定义

$$\boldsymbol{x}^{(r)}=A\boldsymbol{x}^{(r-1)}=\cdots=A^r\boldsymbol{x}^{(0)},\qquad r\geqslant 1,$$

当 $\boldsymbol{x}^{(r)}>\boldsymbol{0}$ 时，并令

$$\bar{\lambda}_r=\max_{1\leqslant i\leqslant n}\frac{x_i^{(r+1)}}{x_i^{(r)}}\quad 与\quad \underline{\lambda}_r=\min_{1\leqslant i\leqslant n}\frac{x_i^{(r+1)}}{x_i^{(r)}},\quad r\geqslant 0.$$

证明：(1) 若 A 不可约，则有 $\underline{\lambda}_0\leqslant\underline{\lambda}_1\leqslant\underline{\lambda}_2\leqslant\cdots\leqslant\rho(A)\leqslant\cdots\leqslant\bar{\lambda}_2\leqslant\bar{\lambda}_1\leqslant\bar{\lambda}_0$，且当且仅当 A 为素矩阵时，对任意 $\boldsymbol{x}^{(0)}>\boldsymbol{0}$ 序列 $\{\bar{\lambda}_r\}$ 与 $\{\underline{\lambda}_r\}$ 均收敛于 $\rho(A)$.（这是 7.1.1 小节定理 17 中关于 $\rho(A)$ 估计的一种改进）

(2) 若 A 满足 $A\boldsymbol{x}>\boldsymbol{0}$，$\forall\,\boldsymbol{x}>\boldsymbol{0}$，则 (1) 中不等式亦成立. 试问：什么时候对任意初始正向量 $\boldsymbol{x}^{(0)}$，这些序列都收敛于 $\rho(A)$？

8. 设 z_j 为复系数多项式 $p_n(z)=z^n-a_1z^{n-1}-a_2z^{n-2}-\cdots-a_n$ 的任一零点，且 r^* 为多项式 $\hat{p}_n(z)=z^n-|a_1|z^{n-1}-|a_2|z^{n-2}-\cdots-|a_n|$ 的最大非负实零点. 证明：$\max\limits_{1\leqslant j\leqslant n}|z_j|\leqslant r^*\leqslant\max\{|a_n|,1+|a_{n-1}|,\cdots,1+|a_1|\}$.

9. 设 $A\in\mathcal{N}_n\,(n\geqslant 2)$ 的每行与每列均不是零向量，且有置换矩阵 P 使得 $P^{\mathrm{T}}AP$ 有 7.1.2 小节 (4) 式的形式，其中块对角线上零块均为方阵. 试证：A 不可约当且仅当 $A_{12}A_{23}\cdots A_{k1}$ 为不可约矩阵.

7.2.2 Perron-Frobenius 定理的进一步推广

7.2.1 小节中介绍的连续性推证方法可以将非负不可约矩阵的 Perron-Frobenius 理论的一些结果推广到一般非负方阵的情形. 但由于极限过程往往不保持严格不等式，因而用连续性推证方法得到的推广结果往往失于精密. 在本小节里，我们将一般非负矩阵的组合结构与非负不可约矩阵的 Perron-Frobenius 理论结合

起来，得到这个理论的较强的推广. 其主要思想是以非负方阵对应于谱半径的代数特征空间中的广义特征向量代替非负不可约矩阵的 Perron 向量. 为此先研究非负方阵的组合结构.

设 $A\in\mathcal{N}_n$，其法式如 7.2.1 小节(2)式所示. 我们约定在法式中若 $m\geqslant 2$，则 A 为可约的；若 $m=1$，则 A 或不可约或为一阶零矩阵. A 的法式的分块法将集合 $\mathcal{N}=\{1,\cdots,n\}$ 划分为 m 个彼此不交的非空子集 $\mathcal{S}_k(1\leqslant k\leqslant m)$. 将每个 $\mathcal{S}_k$ 看成一个等价类，即只有同一类中任意两个不同元素 i 与 j 之间有有向路自 i 到 j 与自 j 到 i. 这时，称这样的不同元素 i 与 j 是**相通的**. 对任意 $\mathcal{S}_i$ 与 $\mathcal{S}_j$，我们称 $\mathcal{S}_i$ **有通路到** $\mathcal{S}_j$，记为 $\mathcal{S}_i\to\mathcal{S}_j$，假如在 A 的有向图 $\Gamma(A)$ 中，有一有向路自 $\mathcal{S}_i$ 内某元素到 $\mathcal{S}_j$ 内某元素. 这样一来，若 A 的法式中 A_{jj} 的脚标集为 $\mathcal{S}_j(1\leqslant j\leqslant m)$，则 $\mathcal{S}_i\to\mathcal{S}_j$ 蕴涵 $i\leqslant j$.

定义 1　设 $A\in\mathcal{N}_n$ 的法式如 7.2.1 小节(2)式所示. 将每个 $\mathcal{S}_k(1\leqslant k\leqslant m)$ 定义为由 $\Gamma(A)$ 相通关系导出一个**等价类**. 假如 $\rho(A_{kk})=\rho(A)$，则等价类 $\mathcal{S}_k$ 称为**基类**；假如其他类没有通路到 $\mathcal{S}_k$，则 $\mathcal{S}_k$ 称为**始类**；假如 $\mathcal{S}_k$ 没有通路到其他类，则 $\mathcal{S}_k$ 称为**终类**.

显然，任意 $A\in\mathcal{N}_n$ 至少有一基类，始类与终类. 特别地在 A 的法式中，$\mathcal{S}_1$ 与 $\mathcal{S}_m$ 分别为始类与终类. 一般地，$\mathcal{S}_k$ 为始类(终类)当且仅当在 A 的法式中 $A_{jk}=O$，$\forall j\neq k(A_{kj}=O,\forall j\neq k)$. 注意，若 $A\in\mathcal{N}_n$ 不可约，则它仅有一个等价类，此唯一的类为基类也为始类与终类，反之亦然(除去 A 为一阶零矩阵的平凡情形).

例 2　设 $A\in\mathcal{N}_{10}$ 为

$$A=\begin{array}{c}\begin{matrix}1&2&3&4&5&6&7&8&9&10\end{matrix}\\ \begin{bmatrix}1&0&8&0&0&8&0&0&0&0\\0&3&6&7&0&0&0&0&0&0\\0&0&2&0&1&0&0&0&0&0\\0&0&0&3&3&0&4&0&0&0\\0&0&0&0&3&5&0&0&0&0\\0&0&0&0&0&3&0&0&0&0\\0&0&0&0&0&0&2&0&0&0\\0&0&0&0&0&0&3&3&0&0\\0&0&0&0&0&0&0&1&1&0\\0&0&0&0&0&0&0&1&1&0\end{bmatrix}\begin{matrix}1\\2\\3\\4\\5\\6\\7\\8\\9\\10\end{matrix}\end{array}.\tag{1}$$

由于 $\rho(A)=3$，故 A 的基类为 $\mathcal{S}_2=\{2\}$，$\mathcal{S}_4=\{4\}$，$\mathcal{S}_5=\{5\}$，$\mathcal{S}_6=\{6\}$ 与 $\mathcal{S}_8=\{8\}$，其余的 $\mathcal{S}_1=\{1\}$，$\mathcal{S}_3=\{3\}$，$\mathcal{S}_7=\{7\}$ 与 $\mathcal{S}_9=\{9,10\}$ 不是基类. 九个等价类之间的相通情况如图 7.3 所示，其中标有圆圈的为基类，方框的为非基类；$\mathcal{S}_i\to\mathcal{S}_j$ 表示 $\mathcal{S}_i$ 类有通路到 $\mathcal{S}_j$ 类. 从图 7.3 中看出，$\mathcal{S}_1$ 与 $\mathcal{S}_2$ 为始类，而 $\mathcal{S}_6$ 与 $\mathcal{S}_9$ 为终类.　□

定义 3　设 $A=(a_{ij})\in\mathcal{N}_n$ 有 m 个等价类 $\mathcal{S}_1,\cdots,\mathcal{S}_m$. 假如 $\mathcal{S}_{j_s}$ 有通路到 $\mathcal{S}_{j_{s+1}}(1\leqslant s\leqslant k-1)$，则称 $\{\mathcal{S}_{j_1},\mathcal{S}_{j_2},\cdots,\mathcal{S}_{j_k}\}$ 为自 $\mathcal{S}_{j_1}$ 到 $\mathcal{S}_{j_k}$ 的一条**链**. 链中含有基类的个数叫作

该链的**长度**. 假如自 $\mathscr{S}_i$ 类到 $\mathscr{S}_j$ 类的最长链的长度为 l，则称 $\mathscr{S}_i$ 类有 l 步的通路到 $\mathscr{S}_j$ 类. 此外，我们称某基类的**高度**为 h，假如以它为终点的最长链的长度为 h.

在例 2 矩阵 A 的等价类相通图中，所有基类的高度见表 7.1.

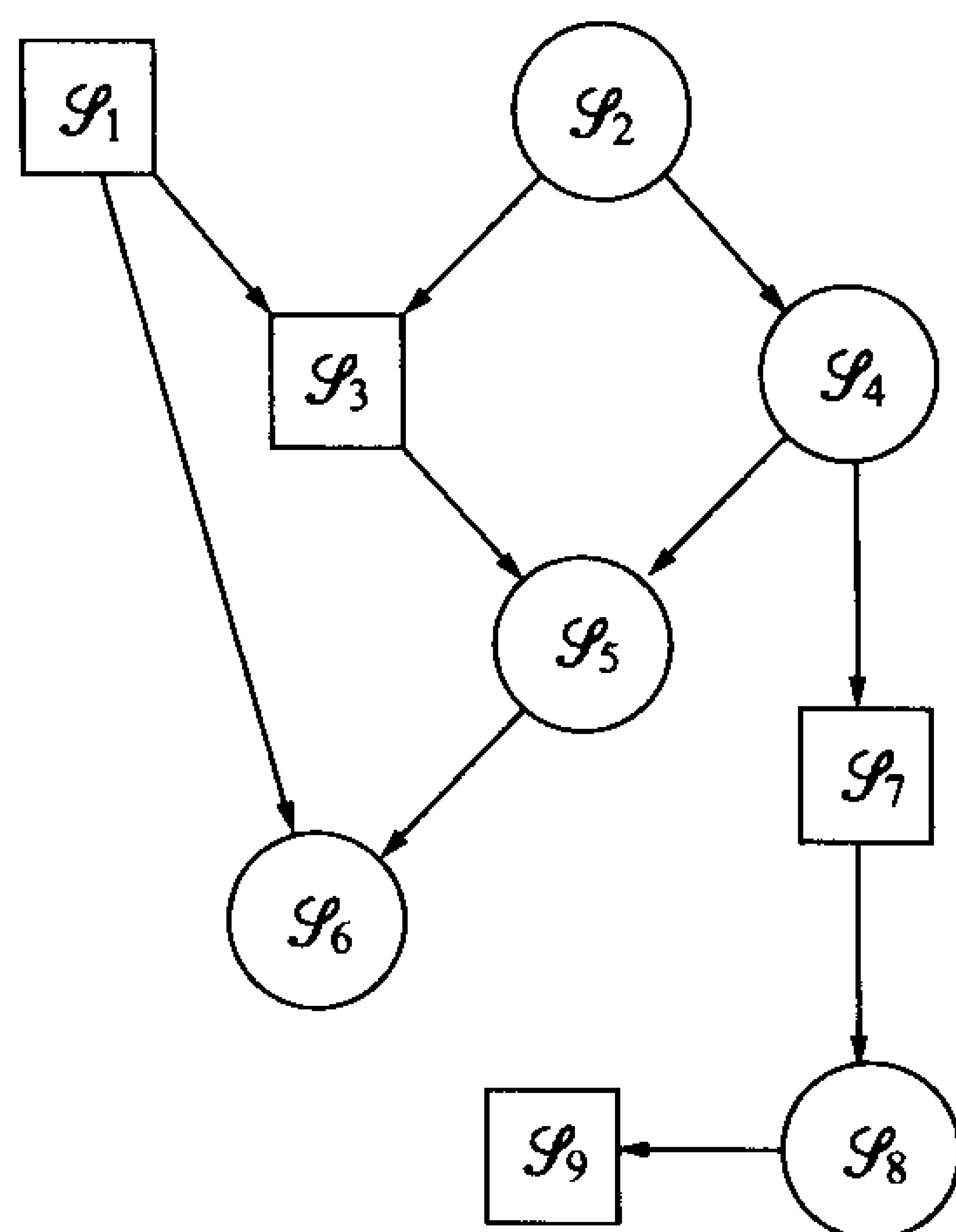


图 7.3　等价类之间相通情况

表 7.1　基类的高度

基　类	高　度
$\mathscr{S}_2$	1
$\mathscr{S}_4$	2
$\mathscr{S}_5$	3
$\mathscr{S}_6$	4
$\mathscr{S}_8$	3

本小节的主要结果为定理 7. 为证明它我们还需要一些预备知识.

引理 4　设 $A\in\mathscr{N}_n$ 有法式如 7.2.1 小节(2)式所示，其等价类为 $\mathscr{S}_1,\cdots,\mathscr{S}_m$. 则 A 有正特征向量对应于 $\rho(A)$ 当且仅当 A 的基类与终类一致.

证明　设有 $\boldsymbol{x}>\boldsymbol{0}$ 满足 $A\boldsymbol{x}=\rho(A)\boldsymbol{x}$. 按 A 的法式的分块法，将 $P^{\mathrm{T}}\boldsymbol{x}$ 分块为 $P^{\mathrm{T}}\boldsymbol{x}=(\boldsymbol{y}_1^{\mathrm{T}},\cdots,\boldsymbol{y}_m^{\mathrm{T}})^{\mathrm{T}}$，其中，各 $\boldsymbol{y}_j$ 为维数与 A_{jj} 阶数相同的正(列)向量. 从 $P^{\mathrm{T}}APP^{\mathrm{T}}\boldsymbol{x}=\rho(A)p^{\mathrm{T}}\boldsymbol{x}$ 推出，

$$A_{kk}\boldsymbol{y}_k+A_{k,k+1}\boldsymbol{y}_{k+1}+\cdots+A_{km}\boldsymbol{y}_m=\rho(A)\boldsymbol{y}_k\qquad(1\leqslant k\leqslant m).\tag{2}$$

假定 $\mathscr{S}_k$ 为基类，则有 $\rho(A_{kk})=\rho(A)$，因而 $\det(\rho(A)I-A_{kk})=0$. 当 A_{kk} 为一阶零矩阵时，$\rho(A_{kk})=\rho(A)=0$，由(2)式得出 $A_{kj}=O,\forall j\neq k$，即 $\mathscr{S}_k$ 为终类. 当 A_{kk} 为不可约矩阵时，$D^{-1}(A_{kk}-\rho(A)I)D$ 也是一个不可约矩阵，这里 D 是以 $\boldsymbol{y}_k$ 分量为对角元素的正对角矩阵. 如若 $\mathscr{S}_k$ 不为终类，由(2)式则有

$$A_{kk}\boldsymbol{y}_k\underset{(\neq)}{\leqslant}\rho(A)\boldsymbol{y}_k.\tag{3}$$

这表明 $D^{-1}(A_{kk}-\rho(A)I)D$ 为不可约对角占优矩阵，于是我们有 $\det(D^{-1}(A_{kk}-\rho(A)I)D)=\det(A_{kk}-\rho(A)I))\neq0$，导出矛盾. 因此，$\mathscr{S}_k$ 必定为终类. 反之，若 $\mathscr{S}_k$ 为终类，则(2)式推出 $A_{kk}\boldsymbol{y}_k=\rho(A)\boldsymbol{y}_k$，因而 $\rho(A_{kk})=\rho(A)$，即 $\mathscr{S}_k$ 为基类. 至此我们证明了条件的必要性.

现设 A 的基类与终类一致. 不失普遍性，假定 $\mathscr{S}_{g+1},\cdots,\mathscr{S}_m$ 为 A 的所有基类(终类)，$\mathscr{S}_1,\cdots,\mathscr{S}_g$ 为余下的等价类. 应用 7.1.1 小节定理 5，A_{kk} 有正特征向量 $\boldsymbol{y}_k$ 对

应于$\rho(A_{kk})=\rho(A)$，$k=g+1,\cdots,m$(注意当A_{kk}为一阶零矩阵时，这个结论亦真．今后往往不另区分A为不可约与一阶零矩阵的情形)．对于$1\leqslant k\leqslant g$，$\rho(A_{kk})<\rho(A)$，于是$(\rho(A)I-A_{kk})^{-1}$存在．此时若A_{kk}为一阶零矩阵，则显然$(\rho(A)I-A_{kk})^{-1}>O$；若A_{kk}非负不可约，则习题7.1.1题11蕴涵$(\rho(A)I-A_{kk})^{-1}>O$．现令

$$\boldsymbol{y}_g=(\rho(A)I-A_{gg})^{-1}\sum_{j=g+1}^{m}A_{gj}\boldsymbol{y}_j.$$

由于上式右端$\sum$中各A_{gj}不全为O，故由前面分析知道$\boldsymbol{y}_g>\boldsymbol{0}$．接着又令

$$\boldsymbol{y}_{g-1}=(\rho(A)I-A_{g-1,g-1})^{-1}\sum_{j=g}^{m}A_{g-1,j}\boldsymbol{y}_j.$$

同理可证明$\boldsymbol{y}_{g-1}>\boldsymbol{0}$．一般可递归地令

$$\boldsymbol{y}_i=(\rho(A)I-A_{ii})^{-1}\sum_{j=i+1}^{m}A_{ij}\boldsymbol{y}_j,\qquad i=g,g-1,\cdots,1,\tag{4}$$

它们都是正向量．现取$\boldsymbol{x}=(\boldsymbol{y}_1^{\mathrm{T}},\cdots,\boldsymbol{y}_m^{\mathrm{T}})^{\mathrm{T}}$．则由上述构造可知，$\boldsymbol{x}>\boldsymbol{0}$且$P^{\mathrm{T}}AP\boldsymbol{x}=\rho(A)\boldsymbol{x}$，于是，$P\boldsymbol{x}$即为我们所需要的对应于$\rho(A)$的$A$的正特征向量．　□

引理4有一个直接的推论，此是习题7.1.1题5的推广结果．它的证明留给读者(见习题1)．

推论5　设$A\in\mathcal{N}_n$有法式如7.2.1小节(2)式所示．则当且仅当A的所有等价类$\mathcal{S}_1,\cdots,\mathcal{S}_m$都是终-基类时，$A$与$A^{\mathrm{T}}$有对应于$\rho(A)$的正特征向量．

现在引入非负方阵A的代数特征空间及其广义特征向量的概念．

定义6　设$A\in\mathcal{N}_n$．满足条件

$$\mathrm{Ker}((\rho(A)I-A)^k)=\mathrm{Ker}((\rho(A)I-A)^{k+1})\tag{5}$$

的最小正整数k，称为A的**度**(degree)，记为$v(A)$．零空间$\mathrm{Ker}((\rho(A)I-A)^{v(A)})$称为$A$的**代数特征空间**，该空间内的元素称为$A$的**广义特征向量**．

说明一下，在有的文献中，非负方阵的度也叫作**指标**(index)或**坡度**(ascent)．实际上，$A\in\mathcal{N}_n$的度即是A的特征值$\rho(A)$的指标(见第三章3.2.1小节)，因而$v(A)$恰为$A\in\mathcal{N}_n$的Jordan正规形式中对应于特征值$\rho(A)$的最大Jordan块的阶数．因此，我们可得出，$\mathrm{Ker}((\rho(A)I-A)^j)\subset\mathrm{Ker}((\rho(A)I-A)^{v(A)})$，$\forall j\geqslant1$．

假如$A\in\mathcal{N}_n$不可约，按7.1.1小节定理14，则有A的度$v(A)=1$，因而这时A的代数特征空间为$\mathrm{Ker}(\rho(A)I-A)$，并且非零的广义特征向量即为对应于$\rho(A)$的A的特征向量．一般地，$A\in\mathcal{N}_n$的代数特征空间的维数恰等于特征值$\rho(A)$的代数重数，即A法式中基类的个数．

定理7(Rothblum[2])　设$A\in\mathcal{N}_n$有法式如7.2.1小节(2)式所示，且A有p个基类$\mathcal{S}_{j_1},\cdots,\mathcal{S}_{j_p}$．则$A$的代数特征空间含有非负向量$\boldsymbol{x}^{(1)},\cdots,\boldsymbol{x}^{(p)}$使得$x_i^{(t)}>0$，当且仅当$i$有通路到基类$\mathcal{S}_{j_t}$，并且，这些非负向量构成$A$的代数特征空间的基．

证明　设基类$\mathcal{S}_{j_t}$的高度为h．我们要证明：$\mathrm{Ker}((\rho(A)I-A)^h)$有一非负向量

$\boldsymbol{x}=(x_1,\cdots,x_n)^{\mathrm{T}}$ 满足 $x_i>0$ 当且仅当 i 有通路到 $\mathscr{S}_{j_t}$ 内某元素. 应用 A 的行列同步变换,我们不妨可以假定 A 呈如下形式:

$$A=\begin{bmatrix} B_h & B_{h,h-1} & \cdots & B_{h1} & B_{h0} \\ & B_{h-1} & & B_{h-1,1} & B_{h-1,0} \\ & & \ddots & \vdots & \vdots \\ & & & B_1 & B_{10} \\ O & & & & B_0 \end{bmatrix}, \tag{6}$$

式中,$B_i(0\leqslant i\leqslant h)$为 A 的子方阵,它对应着 A 的某些等价类,它们有 i 步通路到 $\mathscr{S}_{j_t}$ 类. 这时,若 A_{kk} 为 A 的法式中对应于 $\mathscr{S}_{j_t}$ 类的块阵,则它一定是 B_1 的子阵. 对 $i=1,\cdots,h$,令

$$C_{i+1}=\begin{bmatrix} B_i & B_{i,i-1} & \cdots & B_{i1} \\ & B_{i-1} & \cdots & B_{i-1,1} \\ & & \ddots & \vdots \\ & O & & B_1 \end{bmatrix}, \tag{7}$$

且对 $i=2,\cdots,h$,令

$$D_i=[B_{i,i-1}\cdots B_{i1}]. \tag{8}$$

于是,

$$A=\begin{bmatrix} C_{h+1} & * \\ O & B_0 \end{bmatrix}, \qquad C_{i+1}=\begin{bmatrix} B_i & D_i \\ O & C_i \end{bmatrix}, \qquad \forall\, i\geqslant 2. \tag{9}$$

根据前面的构造,$\rho(B_i)=\rho(A)$,$\forall\, i\geqslant 1$,并且,非负方阵 $B_i(i\geqslant 1)$的基类与它的终类一致. 应用引理 4 有正向量 $\boldsymbol{y}^{(i)}$满足

$$(\rho(A)I-B_i)\boldsymbol{y}^{(i)}=\boldsymbol{0}, \qquad i=1,\cdots,h. \tag{10}$$

同时,由于非负不可约矩阵的度为 1,我们可以推出,凡基类为终类的非负矩阵的度也为 1. 因此,$v(B_i)=1,i=1,\cdots,h$.

现在应用归纳法证明:对 $i=1,\cdots,h$,存在正向量 $\boldsymbol{z}^{(i)}$满足

$$(\rho(A)I-C_{i+1})^i\boldsymbol{z}^{(i)}=\boldsymbol{0}. \tag{11}$$

对 $i=1$,这个结果成立,因为 $C_2=B_1$,此时可选取 $\boldsymbol{z}^{(1)}=\boldsymbol{y}^{(1)}$. 现假定它对 $i-1$ 成立,$1\leqslant i-1<h$,考虑对 i 的情形. 经简单计算看出,

$$\begin{aligned}(\rho(A)I-C_{i+1})^i&=\begin{bmatrix} \rho(A)I-B_i & -D_i \\ O & \rho(A)I-C_i \end{bmatrix}^i \\ &=\begin{bmatrix} (\rho(A)I-B_i)^i & -\sum\limits_{k=0}^{i-1}(\rho(A)I-B_i)^kD_i(\rho(A)I-C_i)^{i-k-1} \\ O & (\rho(A)I-C_i)^i \end{bmatrix}.\end{aligned} \tag{12}$$

按归纳假定与(11),(12)两式,我们有

$$(\rho(A)I-C_{i+1})^{i}\begin{bmatrix} z \\ z^{(i-1)} \end{bmatrix}$$

$$=\begin{bmatrix} (\rho(A)I-B_i)^{i}z-\sum_{k=1}^{i-1}(\rho(A)I-B_i)^{k}D_i(\rho(A)I-C_i)^{i-k-1}z^{(i-1)} \\ \mathbf{0} \end{bmatrix}$$

对任意 z 成立. 因为 $v(B_i)=1(1\leqslant i\leqslant h)$,由(5)式看出,$(\rho(A)I-B_i)^{k}$ 的列空间与 k 的选取无关,所以$\sum_{k=1}^{i-1}(\rho(A)I-B_i)^{k}D_i(\rho(A)I-C_i)^{i-k-1}z^{(i-1)}$属于$(\rho(A)I-B_i)^{i}$的列空间. 于是,存在某实向量 z 使得

$$(\rho(A)I-C_{i+1})^{i}\begin{bmatrix} z \\ z^{(i-1)} \end{bmatrix}=\mathbf{0}. \tag{13}$$

选择正实数 λ 满足 $\lambda y^{(i)}+z>\mathbf{0}$,并令

$$z^{(i)}=\begin{bmatrix} \lambda y^{(i)}+z \\ z^{(i-1)} \end{bmatrix}.$$

从(10),(12)与(13)诸式即推出(11)式. 现选取

$$x=\begin{bmatrix} z^{(h)} \\ \mathbf{0} \end{bmatrix},$$

从(9),(11)式以及 $z^{(h)}>\mathbf{0}$ 的事实我们看出,$x\geqslant\mathbf{0}(\neq\mathbf{0})$且

$$(\rho(A)I-A)^{h}x=\begin{bmatrix} (\rho(A)I-C_{h+1})^{h}z^{(h)} \\ \mathbf{0} \end{bmatrix}=\mathbf{0},$$

同时,当且仅当 i 有通路到基类 $\mathscr{S}_{j_t}$ 时,x 的第 i 个分量为正数:$x_i>0$. 因此,更有 $(\rho(A)I-A)^{v(A)}x=\mathbf{0}$, 即 x 为 A 的广义特征向量. 现设 $x^{(t)}\geqslant\mathbf{0}(1\leqslant t\leqslant p)$ 为 A 的广义特征向量,满足条件:当且仅当 i 有通路到基类 $\mathscr{S}_{j_t}$ 时,$x_i^{(t)}>0(1\leqslant t\leqslant p)$. 余下我们要证明:这样的 $x^{(1)},x^{(2)},\cdots,x^{(p)}$ 组成 A 的代数特征空间的基. 但这个空间的维数等于 A 的法式中基类个数 p,因此我们只需证明 $x^{(1)},x^{(2)},\cdots,x^{(p)}$ 线性无关. 假定 A 的基类 $\mathscr{S}_{j_1},\cdots,\mathscr{S}_{j_p}$ 已按 $j_1<j_2<\cdots<j_p$ 顺序排列,且设$\sum_{i=1}^{p}\alpha_i x^{(i)}=\mathbf{0}$. 令 $x^{(t)}(s)$ 为 $x^{(t)}$ 的子向量,它对应于基类 $\mathscr{S}_{j_s}$. 这时,$x^{(t)}(p)=\mathbf{0},\forall t<p$. 但由于 $x^{(p)}(p)>\mathbf{0}$,故有 $\alpha_p=0$. 用类似方法可证,$\alpha_{p-1}=\cdots=\alpha_1=0$,即有 $x^{(1)},\cdots,x^{(p)}$ 线性无关. □

上述定理证明的关键在于将 $A\in\mathscr{N}_n$ 的法式变换为形式(6). 为了便于理解这个证明,我们考虑例 2 中的矩阵 $A\in\mathscr{N}_{10}$.

例 8　对于例 2 中的矩阵 $A\in\mathscr{N}_{10}$,有 $\rho(A)=3,v(A)=4$,且 $\mathscr{S}_2,\mathscr{S}_4,\mathscr{S}_5,\mathscr{S}_6$ 与 $\mathscr{S}_8$ 为基类. 现取 $j_4=6$. 因基类 $\mathscr{S}_6$ 高度 $h=4$(但一般地 h 不等于 $v(A)$,如对 $\mathscr{S}_4,h=$

$2 < v(A)$)，所以(6)式矩阵中，按图 7.3 我们有

$$B_0 = \underset{7\ \ 8\ \ 9\ \ 10}{\begin{bmatrix} 0 & 2 & 0 & 0 \\ 0 & 3 & 3 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}}, \quad B_1 = \underset{6}{[3]}, \quad B_2 = \underset{1\ \ 3\ \ 5}{\begin{bmatrix} 1 & 8 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 3 \end{bmatrix}},$$

$$B_3 = \underset{4}{[3]}, \qquad B_4 = \underset{2}{[3]}.$$

因此，(6)式右端矩阵为

$$P^{\mathrm{T}}AP = \widetilde{A} = \begin{array}{c} \begin{array}{cccccccccc} 2 & 4 & 1 & 3 & 5 & 6 & 7 & 8 & 9 & 10 \end{array} \\ \left[\begin{array}{c|c|ccc|c|cccc} 3 & 7 & 0 & 6 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 3 & 0 & 0 & 3 & 0 & 4 & 0 & 0 & 0 \\ \hline 0 & 0 & 1 & 8 & 0 & 8 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 5 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{array}\right] \begin{array}{c} 2 \\ 4 \\ 1 \\ 3 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \end{array} \end{array}$$

(注意，不同于 A 的法式，这里的 $B_0,\cdots,B_4$ 虽为方阵，但它们如 B_0 与 B_2 可以为非零可约矩阵)在 B_1,B_2,B_3 与 B_4 中，它们的某等价类为 A 的基类，因而也是 $B_1,\cdots,B_4$ 的基类. 并且 $B_i(1\leqslant i\leqslant 4)$中的基类与终类一致. 满足

$$(\rho(A)I - B_i)\boldsymbol{y}^{(i)} = \mathbf{0}, \qquad i = 1,2,3,4$$

的正向量可取为 $\boldsymbol{y}^{(1)}=(1)$，$\boldsymbol{y}^{(2)}=(4,1,1)^{\mathrm{T}}$，$\boldsymbol{y}^{(3)}=\boldsymbol{y}^{(4)}=(1)$. 考虑矩阵$(\rho(A)I-C_{i+1})^i(1\leqslant i\leqslant 4)$. 对 $i=1$，取 $\boldsymbol{z}^{(1)}=\boldsymbol{y}^{(1)}=(1)$，对 $i=2$，由于

$$(\rho(A)I - C_3)^2 = \begin{bmatrix} 2 & -8 & 0 & -8 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & -5 \\ 0 & 0 & 0 & 0 \end{bmatrix}^2 = \begin{bmatrix} 4 & -24 & 8 & -16 \\ 0 & 1 & -1 & 5 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

因此可取$\begin{bmatrix} \boldsymbol{z} \\ \boldsymbol{z}^{(1)} \end{bmatrix}=(-2,1,6,1)^{\mathrm{T}}$，用 $\boldsymbol{y}^{(2)}+\boldsymbol{z}=(2,2,7)^{\mathrm{T}}$ 代替 $\boldsymbol{z}=(-2,1,6)^{\mathrm{T}}$，并令 $\boldsymbol{z}^{(2)}=(2,2,7,1)^{\mathrm{T}}$，它满足$(\rho(A)I-C_3)^2\boldsymbol{z}^{(2)}=\mathbf{0}$. 对 $i=3$，

$$(\rho(A)I-C_4)^3=\begin{bmatrix}0&0&0&-3&0\\0&2&-8&0&-8\\0&0&1&-1&0\\0&0&0&0&-5\\0&0&0&0&0\end{bmatrix}^3=\begin{bmatrix}0&0&0&0&0\\0&8&-56&24&-72\\0&0&1&-1&5\\0&0&0&0&0\\0&0&0&0&0\end{bmatrix},$$

因此可取 $\boldsymbol{z}^{(3)}=(1,2,2,7,1)^{\mathrm{T}}$，它满足$(\rho(A)I-C_4)^3\boldsymbol{z}^{(3)}=\mathbf{0}$. 同理，对 $i=4$ 可取 $\boldsymbol{z}^{(4)}=(1,1,2,2,7,1)^{\mathrm{T}}$，它满足$(\rho(A)I-C_5)^4\boldsymbol{z}^{(4)}=\mathbf{0}$. 最后，取 $\tilde{\boldsymbol{x}}=(1,1,2,2,7,1,0,0,0,0)^{\mathrm{T}}$，它便满足$(\rho(A)I-\widetilde{A})^4\tilde{\boldsymbol{x}}=\mathbf{0}$. 令 $\boldsymbol{x}=P\tilde{\boldsymbol{x}}=(2,1,2,1,7,1,0,0,0,0)^{\mathrm{T}}$. 它是 A 的非负广义特征向量(即前一定理中的 $\boldsymbol{x}^{(4)}$)，当且仅当 i 有通路到基类 $\mathscr{S}_6$ 时，$x_i>0$.

假如我们分别再从 $\mathscr{S}_2,\mathscr{S}_4,\mathscr{S}_5$ 与 $\mathscr{S}_8$ 入手，也可以得到 A 的四个广义特征向量(即为前一定理中的 $\boldsymbol{x}^{(1)},\boldsymbol{x}^{(2)},\boldsymbol{x}^{(3)}$ 与 $\boldsymbol{x}^{(5)}$)，它们也满足定理 7 中的条件. 这些广义特征向量 $\boldsymbol{x}^{(1)},\cdots,\boldsymbol{x}^{(5)}$ 组成 $\mathrm{Ker}(3I-A)^4$ 的基(见习题 6). □

我们指出，定理 7 给出 7.1.1 小节定理 5 在一般非负方阵情形的推广，因为如将定理 7 限定在非负不可约矩阵假定之下，则 A 只有一个基类，且 A 的代数特征空间为 $\mathrm{Ker}(\rho(A)I-A)$，该定理的结论便退化为 7.1.1 小节定理 5 的基本结果. 另外，定理 7 的证明是构造性的，它提供寻求非负方阵广义特征向量的一种办法.

关于 Perron-Frobenius 理论在一般非负方阵情形的推广还有两个补充结论. 限于篇幅我们仅列出结果而略去证明(读者可参考文献[2]).

定理 9 设 $A\in\mathscr{N}_n$. 则 A 的度 $v(A)$ 等于它的最长链的长度.

例如，对例 2 中非负矩阵 A 的最长链的长度为 4，于是，$v(A)=4$. 一般地，从定理 9 看出，$A\in\mathscr{N}_n$ 的任意基类的高度$\leqslant v(A)\leqslant A$ 的基类的个数，特别有 $v(A)\leqslant n$.

定理 10 设 $A\in\mathscr{N}_n$ 的度为 $v(A)$. 则在 A 的所有广义特征向量之中存在某向量 $\boldsymbol{x}$，它具有最多的正分量，且对 $0\leqslant k\leqslant v(A)-1$，$(A-\rho(A)I)^k\boldsymbol{x}\underset{(\neq)}{\geqslant}\mathbf{0}$. 此外，当且仅当 i 有至少 $k+1$ 步的通路到 A 某基类时，$((A-\rho(A)I)^k\boldsymbol{x})_i>0$.

我们指出，当 $A\in\mathscr{N}_n$ 不可约时，$v(A)=1$，上述定理中的 $\boldsymbol{x}$ 即为 A 的 Perron 向量.

习题 7.2.2

1. 证明推论 5.
2. 继续研究 7.2.1 小节(3)式中矩阵. 应用它来验证定理 7 的结论.
3. 设非负矩阵 A 为

$$A=\mathrm{diag}[J_1,J_2,J_3],$$

其中，

$$J_1 = \begin{bmatrix} 3 & 1 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{bmatrix}, J_2 = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}, J_3 = \begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}.$$

试求 $v(A)$，并证明：$\mathrm{Ker}(\rho(A)I-A)^k=\mathrm{Ker}(\rho(A)I-A)^{v(A)}$，$\forall k\geqslant v(A)$. 这个结果对一般非负方阵也成立吗？

4. 设 $A\in\mathcal{N}_n$ 有法式如 7.2.1 小节(2)式所示，且其等价类为 $\mathscr{S}_1,\cdots,\mathscr{S}_m$. 试证：当且仅当 A 的终类为基类时，A 有正的广义特征向量；并且，当且仅当任一基类没有通路到其他基类时，$v(A)=1$.

5. 证明：若 $A\in\mathcal{N}_n$ 如题 4 所设，则当且仅当 A 的基类为终类时，存在 $\boldsymbol{x}>\boldsymbol{0}$ 使得 $(\rho(A)I-A)\boldsymbol{x}\geqslant\boldsymbol{0}$；当且仅当 A 的基类为始-终类时，存在 n 阶正对角矩阵 D 使得 $AD+DA^{\mathrm{T}}$ 为正半定的.

6. 分别从 $\mathscr{S}_2,\mathscr{S}_4,\mathscr{S}_5$ 与 $\mathscr{S}_8$ 入手，继续完成例 8 的计算，求出相应的 A 的广义特征向量. 验证一下，这四个向量连同例 8 中已求出的广义特征向量构成 A 的代数特征空间的基.（建议：可与他人共同完成有关计算）

7.3 随机矩阵与双随机矩阵

矩阵 $A=(a_{ij})\in\mathcal{N}_n$ 叫作**随机矩阵**(stochastic matrix)，假如它的所有行元素之和 $\sum_{j=1}^{n}a_{ij}=1(1\leqslant i\leqslant n)$. 矩阵 $A=(a_{ij})\in\mathcal{N}_n$ 叫作**双随机矩阵**(doubly stochastic matrix)，假如 A 与 A^{T} 皆为随机矩阵. 这两种特殊的非负矩阵具有重要的应用价值，尤其在概率论的有限 Markov 过程理论中起着基本的作用.

7.3.1 随机矩阵与有限齐次 Markov 链

随机矩阵又称行随机矩阵. 先考虑它的一个背景材料. 设某个过程有 n 个可能状态(state) $s_1,\cdots,s_n$，且 $t_0,t_1,t_2,\cdots$ 是时刻的一个序列. 假如在每一时刻过程可以处状态 $s_1,\cdots,s_n$ 之一，而只在 $t_0,t_1,t_2,\cdots$ 诸时刻改变其状态，并且，在时刻 t_k 过程从状态 s_i 转移到下一个状态 s_j 的概率 t_{ij} 只依赖于这两个状态，而与时刻 t_k 的下标 k 无关，则称该过程为有限**齐次 Markov 过程**. 这时，n 阶矩阵 $T=(t_{ij})$ 满足 $t_{ij}\geqslant 0$，$\forall i,j$，与 $\sum_{j=1}^{n}t_{ij}=1$，$\forall i$，即 T 为 n 阶随机矩阵，我们称之为该过程的**转移矩阵**. 反之，每一个随机矩阵可作为某个有限齐次 Markov 过程的转移矩阵.

随机矩阵也常出现在数理经济学与运筹学的各种模型问题里.

所有 n 阶随机矩阵的集合显然为 $\boldsymbol{M}_n(\mathbb{R})$ 内的凸体(即紧凸集)，它具有一些简单而又重要的性质，譬如，同阶随机矩阵之积仍为随机矩阵；随机矩阵的谱半径等于 1(按 7.2.1 小节定理 6)，等等. 此外，按定义可立即推出以下结论.

定理 1 设 $A\in\mathcal{N}_n$. 则 A 为随机矩阵当且仅当 $\boldsymbol{e}=(1,\cdots,1)^{\mathrm{T}}\in\mathbb{R}^n$ 为 A 对应

于特征值 1 的特征向量，即 $Ae=e$.

此定理表明，任意随机矩阵的谱半径 1 为其特征值，它对应有正特征向量 e. 于是，按 7.2.2 小节引理 4，任意随机矩阵的法式中，基类与终类相重合. 因此，随机矩阵 A 的度或指标为 1：$v(A)=1$（习题 1).

下面定理告诉我们，具有正谱半径与对应的正特征向量的非负矩阵（这样非负矩阵的特征见 7.2.2 小节引理 4）与随机矩阵之间存在着密切的关系.

定理 2　设 $A\in\mathcal{N}_n$ 的谱半径 $\rho(A)>0$，且有 $\mathbf{y}>\mathbf{0}$ 使得 $A\mathbf{y}=\rho(A)\mathbf{y}$. 则存在正对角矩阵 D 使得 $D^{-1}AD/\rho(A)$ 为随机矩阵.

证明　令 $\mathbf{y}=(y_1,\cdots,y_n)^{\mathrm{T}}>\mathbf{0}$ 与 $D=\mathrm{diag}[y_1,\cdots,y_n]$. 则 $D^{-1}AD/\rho(A)\equiv P\in\mathcal{N}_n$ 满足 $Pe=PD^{-1}\mathbf{y}=D^{-1}A\mathbf{y}/\rho(A)=D^{-1}\mathbf{y}=e$. 因此按定理 1，$P$ 为随机矩阵. □

在实际应用中，经常需要考察随机矩阵 A 的幂序列 $\{A^r\}_{r=1}^{\infty}$ 的收敛性. 由于 $\rho(A)=1$ 且 $v(A)=1$（习题 1)，故我们有如下结果.

定理 3　设 A 为 n 阶不可约随机矩阵. 则 $\lim\limits_{r\to\infty}A^r$ 存在当且仅当 A 为素矩阵.

证明　应用第二章 2.2.2 小节定理 7 即可. □

我们指出，上述定理中不可约性假定可以去掉，但其中“A 为素矩阵”的条件应改换为“A 的非 1 特征值的模皆小于 1”. 事实上，对于随机矩阵 A，不管它是否不可约的，特征值 1 只有线性初等因子，按照 A 的法式结构，A 的模为 1 的特征值也只有线性初等因子（见习题 1).

此外，如设 n 阶随机矩阵 A 的不同特征值为 $\lambda_1,\cdots,\lambda_s$，其中 $\lambda_1=1,\lambda_2,\cdots,\lambda_k$ 的模为 $1,\lambda_{k+1},\cdots,\lambda_s$ 的模小于 1，用 f_{lj} 表示 $f(\lambda)=\lambda^r$ 在 $\lambda_l\in\sigma(A)$ 处的 j 阶导数值，则 $f(A)=A^r$ 有谱分解（见第三章 3.2.2 小节）：

$$A^r=E_{10}+\sum_{l=2}^{k}f_{l0}E_{l0}+\sum_{l=k+1}^{s}\sum_{j=0}^{m_l-1}f_{lj}E_{lj}. \tag{1}$$

这时，假如 A 还为素矩阵（或者 A 的非 1 特征值的模皆小于 1)，那么，$k=1$，因而 (1) 式右端第二项不出现，且由于 $|\lambda_l|<1,\forall\, l\geqslant 2$，故当 $r\to\infty$ 时，(1) 式右端第三项趋于 n 阶零矩阵. 因此，$\lim\limits_{r\to\infty}A^r=E_{10}$（其实此结论也可以从习题 7.1.2 题 8 直接得出，因为这时 $\rho(A)=1$).

还要指出，n 阶随机矩阵集合有一个结构特性. 前面讲过此集合为 $\mathbf{M}_n(\mathbb{R})$ 中的凸体，按著名的 Krein-Milman 定理，**一个凸体 $\mathscr{C}$ 为它的极端点**（extreme point）**的凸包**. 这些极端点称为凸体 $\mathscr{C}$ 的**顶点**. 这里，$X_0\in\mathscr{C}$ 叫作 $\mathscr{C}$ 的极端点，假如 X_0 不能表示为 $X_0=\sum\limits_{j=1}^{s}\alpha_jY_j$，其中 $X_0\neq Y_j\in\mathscr{C},\alpha_j\geqslant 0,\sum\limits_{j=1}^{s}\alpha_j=1$，即 X_0 不是 $\mathscr{C}$ 内其他点的凸组合. n 阶随机矩阵集合有 n^n 个顶点，它们是每行只有一个非零元素 1 的 n 阶矩阵. 因为容易看出，这样的矩阵不能表示为其他 n^n-1 个同类型矩阵的凸组合.

显然，所有 n 阶置换矩阵(共有 $n!$ 个)皆为这个凸体的顶点. 因此，由 Krein-Milman 定理，任意一个 n 阶随机矩阵可以表示为这些 n^n 个顶点的凸组合. 对这方面有兴趣的读者可参阅文献[3].

现在考虑随机矩阵在齐次 Markov 链中的应用. 用 $\mathscr{M}=(T,\boldsymbol{\pi}^0)$ 表示某个有限齐次 Markov 链，这意味着它是以 $T=(t_{ij})\in \boldsymbol{M}_n(\mathbb{R})$ 为转移矩阵，且有**初始(概率)分布向量** $\boldsymbol{\pi}^0=(\pi_1^0,\cdots,\pi_n^0)$ 的 Markov 过程，这里 π_j^0 表示过程初始处在状态 s_j 的概率 $(1\leqslant j\leqslant n)$. 更一般地，用 $\boldsymbol{\pi}^k=(\pi_1^k,\cdots,\pi_n^k)$ 表示**第 k 个(概率)分布向量**，其中，π_j^k 为过程在 k 步后处在状态 s_j 的概率 $(1\leqslant j\leqslant n)$. 显然，$\pi_j^k\geqslant 0,\forall j,k$ 且 $\sum\limits_{j=1}^{n}\pi_j^k=1,\forall k$.

引理 4　设 $\mathscr{M}=(T,\boldsymbol{\pi}^0)$ 为齐次 Markov 链. 则有

$$\boldsymbol{\pi}^k=\boldsymbol{\pi}^{k-1}T=\boldsymbol{\pi}^0T^k,\qquad \forall k\geqslant 1. \tag{2}$$

证明　对任意给定的 $k\geqslant 1$，由于过程移到状态 s_j 的概率只依赖于前一状态，故过程在 $k-1$ 步后处在状态 s_m 且接着转移到状态 s_j 的概率为 $\pi_m^{k-1}t_{mj}$. 因此，过程在 k 步后处在状态 s_j 的概率应是这些概率之和，即

$$\pi_j^k=\sum_{m=1}^{n}\pi_m^{k-1}t_{mj}=(\boldsymbol{\pi}^{k-1}T)_j,$$

由此直接推出(2)式. □

假如 $\boldsymbol{\pi}^k=\boldsymbol{\pi}^0,\forall k\geqslant 1$，则我们称该 Markov 链的初始分布向量 $\boldsymbol{\pi}^0$ 是**平稳的**(stationary). 这样的向量称为**平稳概率分布向量**.

研究 Markov 链中一个重要问题是讨论当 $k\to\infty$ 时，$\boldsymbol{\pi}^k$ 的变化趋势. 按(2)式这主要取决于当 $k\to\infty$ 时，T^k 的变化趋势. 考虑一个实例.

例 5　甲乙二人做游戏：从装有两个红球与一个黄球的口袋里甲每次取一个球，若为黄球，甲付乙 1 元钱；若为红球，乙付甲 1 元钱. 开始，甲有 1 元钱，而乙有 3 元钱. 当二人中一人输了所有钱时游戏终止. 试求每人取胜的概率与此游戏能进行十次的概率.

以甲为主体对象用他的钱数描述各种可能的状态：状态 s_1 表示甲没有钱，状态 s_2 表示甲有 1 元钱，状态 s_3 表示甲有 2 元钱，状态 s_4 表示甲有 3 元钱，状态 s_5 表示甲有 4 元钱. 显然，该过程的转移矩阵 $\widetilde{T}$ 为

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0\\ \frac{1}{3} & 0 & \frac{2}{3} & 0 & 0\\ 0 & \frac{1}{3} & 0 & \frac{2}{3} & 0\\ 0 & 0 & \frac{1}{3} & 0 & \frac{2}{3}\\ 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

假如重新将状态排序为$\{s_1,s_2,s_3,s_4,s_5\}=\{0,4,1,2,3\}$，则过程的转移矩阵 T 变为

$$T=\left[\begin{array}{cc:ccc}1&0&0&0&0\\0&1&0&0&0\\ \hdashline \frac{1}{3}&0&0&\frac{2}{3}&0\\0&0&\frac{1}{3}&0&\frac{2}{3}\\0&\frac{2}{3}&0&\frac{1}{3}&0\end{array}\right]\equiv\begin{bmatrix}I&O\\B&C\end{bmatrix}. \tag{3}$$

我们有

$$T^k=\begin{bmatrix}I&O\\ \sum_{i=0}^{k-1}C^iB&C^k\end{bmatrix},\quad \forall k\geqslant 1, \tag{4}$$

且因 $\rho(C)<1$，所以有$\lim_{k\to\infty}C^k=O$ 与

$$\lim_{k\to\infty}\sum_{i=0}^{k-1}C^iB=(I-C)^{-1}B=\frac{1}{15}\begin{bmatrix}7&8\\3&12\\1&14\end{bmatrix}.$$

因此，

$$\lim_{k\to\infty}T^k=\begin{bmatrix}1&0&0&0&0\\0&1&0&0&0\\ \frac{7}{15}&\frac{8}{15}&0&0&0\\ \frac{3}{15}&\frac{12}{15}&0&0&0\\ \frac{1}{15}&\frac{14}{15}&0&0&0\end{bmatrix}. \tag{5}$$

按假定初始概率分布向量为 $\pi^0=(0,0,1,0,0)$. 应用引理 4 我们有

$$\pi\equiv\lim_{k\to\infty}\pi^k=\pi^0\lim_{k\to\infty}T^k=\left(\frac{7}{15},\frac{8}{15},0,0,0\right),$$

这是链 $\mathscr{M}=(T,\pi^0)$的平稳概率分布向量. 于是，甲最终处于状态 s_1(即甲输)的概率为 7/15，甲最终处于状态 s_2(即乙输)的概率为 8/15，并且，游戏无限进行下去的概率为零. 类似地从计算 T^9 可推出，进行九轮后，甲处于状态 s_3 或 s_5 的概率为零，处于状态 s_4 的概率为$(2/3)^9$. 因此，能进行十次游戏的概率为$(2/3)^9\approx0.026$. □

与概率分布向量 π^k 变化趋势有关的问题是如何区分一个过程中的状态以及不同的过程. 应用本章前面的知识，这两个问题都比较容易地解决.

设某齐次 Markov 链的状态为 $s_1,\cdots,s_n$，且转移矩阵 T 有形如 7.2.1 小节(2)式的法式(但其中 A_{ij} 改为 T_{ij}). 我们说，状态 s_i 有**通路**到 s_j，假如在 T 的有向图 $\Gamma(T)$ 中顶点 i 有通路到顶点 j；状态 s_i 与 s_j **相通**，假如 s_i 有通路到 s_j 且 s_j 有通路到 s_i. 按状态相通的关系可将状态分为等价类. 某类称为**终类**，若 T 法式中对应的对角块为终类. 若终类只含一个状态 s_i，则此状态称为**吸收态**(absorbing state)，这时等价于 $t_{ii}=1$.

从 T 的法式看出，每个齐次 Markov 链至少有一个终类，且此链只由一个(终)类组成当且仅当 $T\in\mathcal{N}_n$ 不可约.

一个状态 s_i 叫作**瞬态**，若 s_i 有通路到某个 s_j，但此 s_j 没有通路到 s_i，否则 s_i 叫作**遍历态**. 因此，s_i 为遍历态当且仅当 s_i 有通路到 s_j 蕴涵 s_j 有通路到 s_i. 含有一个瞬态的类叫作**瞬态类**(因而该类中所有状态为瞬态)，否则叫作**遍历类**(因而该类中所有状态为遍历态). 因此，某类为遍历的当且仅当它是终类.

对于例 5 中五个状态，从(3)式看出，s_1 与 s_2 为吸收态(它们分别表示若甲没有钱与有 4 元钱，则游戏终止). 这时，$\{s_1\}$ 与 $\{s_2\}$ 为遍历类. 但是，状态 s_3,s_4 与 s_5 皆为瞬态(分别对应于甲有 1,2 与 3 元钱)，因而 $\{s_3,s_4,s_5\}$ 组成该链的瞬态类.

一个齐次 Markov 链叫作**遍历链(吸收链)**，如果它由单个遍历类组成(如果其中每一个遍历类只由单个吸收态组成). 该链叫作**规则的**，如果在某固定步 k，链的任意状态移到每个状态的概率均为正数. 该链叫作**周期的**，假如它是遍历的但非规则的.

显然，每个规则 Markov 链为遍历的，但反之不真. 例如，以 $T=\begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}$ 为转移矩阵的链为遍历的但不是规则的，因而为周期的.

应用转移矩阵 T 的性质可以区别 Markov 链的类型.

定理 6　设 T 为某 Markov 链的转移矩阵. 则

(1) 该链为遍历的当且仅当 T 为不可约矩阵.

(2) 该链为规则的当且仅当 T 为素矩阵.

(3) 该链为周期的当且仅当 T 为不可约循环矩阵.

证明　(1) T 不可约当且仅当 $\Gamma(T)$ 强连接，后者等价于该链的每个状态有通路到其他状态，这表明该链为遍历的.

(2) 该链为规则的当且仅当 T 的某个正整数幂为正矩阵，后者等价于 T 为素矩阵.

(3) 该链为周期的当且仅当 T 不可约但不是素矩阵，后者等价于 T 为不可约循环矩阵. □

应用转移矩阵还可以研究平稳概率分布向量的存在性与唯一性问题.

定理 7　每个有限齐次 Markov 链有平稳概率分布向量.

证明　设 T 为某链的转移矩阵. 因 $\rho(T)=1$，按 7.2.1 小节定理 1，有非零行向量 $\boldsymbol{x}\geqslant\boldsymbol{0}$ 使得 $\boldsymbol{x}T=\boldsymbol{x}$. 令 $\boldsymbol{\pi}=\boldsymbol{x}/\sum_{i=1}^{n}x_i$，则有 $\boldsymbol{\pi}T=\boldsymbol{\pi},\pi_1+\cdots+\pi_n=1$，因而 $\boldsymbol{\pi}T^k=\boldsymbol{\pi}^k=\boldsymbol{\pi},\forall k\geqslant 1$，即 $\boldsymbol{\pi}$ 为此链的平稳概率分布向量. □

若 T 为随机矩阵，那么由前面分析知道，它的终类与基类重合，因而 T 的度 $v(T)=1$(见习题 7.2.2 题 4 或本小节习题题 1). 如限定 T 不可约，则有

定理 8　设 $T\in\mathcal{N}_n$ 为随机矩阵. 则 T 为某个遍历的 Markov 链的转移矩阵当且仅当下面两个条件均成立：

(1) 1 为 T 的单重特征值.

(2) 不计正倍数，存在唯一的正行向量 $\boldsymbol{x}$ 使得 $\boldsymbol{x}T=\boldsymbol{x}$.

证明　由定理 6，T 为某遍历链的转移矩阵的结论等价于 T 为不可约的. 按习题 7.1.1 题 5，$T\in\mathcal{N}_n$ 不可约当且仅当 $\rho(T)=1$ 为 T 的单重特征值且 T 与 T^{T} 有正特征向量对应于特征值 1. 但按定理 1，$T\boldsymbol{e}=\boldsymbol{e},\boldsymbol{e}>\boldsymbol{0}$. 再按 7.1.1 小节定理 13，$T^{\mathrm{T}}$ 的正特征向量(不计正倍数)为唯一的. □

推论 9　一个齐次 Markov 链为遍历的当且仅当它有唯一的正平稳概率分布向量.

习题 7.3.1

1. 试证：任意随机矩阵 T 的特征值 1 与模为 1 的特征值均只有线性初等因子. 特别地，T 的度 $v(T)=1$.

2. 对 $n=2$ 的情形，直接验证：所有二阶随机矩阵的集合为它的 2^2 个顶点的凸包.

3. 设某少数民族现有 1800 人，居住在 A,B,C 三个部落的人数分别为 200，600，1000. 假定每年每个部落的所有人分为相等的两半分别迁往其他两个部落，试问：一年后、二年后以及无限长久之后，该民族在 A,B,C 三部落的人口分布情况.

4. 设 $T\in\mathcal{N}_n$ 为随机矩阵. 证明：T 为某个规则的 Markov 链的转移矩阵当且仅当 $\delta(T)=\max\{|\lambda|:\lambda\in\sigma(T),\lambda\neq\rho(T)\}<1$，且定理 8 中(1)与(2)成立.

5. 设 $T\in\mathcal{N}_n$ 为随机矩阵. 证明：T 为某个周期的 Markov 链的转移矩阵当且仅当定理 8 中(1)与(2)成立，并且，存在置换矩阵 P 使得

$$P^{\mathrm{T}}TP=\begin{bmatrix} O & T_1 & & \cdots & O \\ O & O & T_2 & \cdots & O \\ \vdots & & \ddots & \ddots & \vdots \\ & & & & T_{h-1} \\ T_h & O & \cdots & & O \end{bmatrix},$$

式中，$h>1$ 且零对角块为方阵.

6. 设非负矩阵 A 的谱半径 $\rho(A)$ 对应有正的特征向量. 试证：A 的模为 $\rho(A)$ 的所有特征值均只有线性初等因子.

7.3.2 双随机矩阵

在第六章 6.3.2 小节定理 10 证明中，我们已经接触到双随机矩阵，并知道由 n 阶酉矩阵 $U=(u_{ij})$ 导出的矩阵 $A=(|u_{ij}|^2)$ 必为 n 阶双随机矩阵. 此外，任何置换矩阵是双随机矩阵.

双随机矩阵类显然地为 7.3.1 小节中随机矩阵类的子类，因而它具有 7.3.1 小节中随机矩阵的所有概念与有关性质. 于是，如下结果是明显的(其中命题(3)的证明可应用习题 7.3.1 题 5 的后半部结果).

定理 1　(1) 设 $A\in\mathcal{N}_n$ 为双随机矩阵. 则 $\rho(A)=1$ 为 A 的特征值，且 $e=(1,\cdots,1)^{\mathrm{T}}$ 为 A 与 A^{T} 对应于特征值 1 的正特征向量.

(2) 两个 n 阶双随机矩阵之积仍为双随机矩阵.

(3) 若 $A\in\mathcal{N}_n$ 为不可约的双随机矩阵，则它的循环指标 k 整除 n，且存在置换矩阵 P 与 Q 使得

$$PAQ=\mathrm{diag}[A_1,A_2,\cdots,A_k],$$

其中，每个 A_j 为 n/k 阶的双随机矩阵.

(4) 若 $A\in\mathcal{N}_n$ 为可约双随机矩阵，且有置换矩阵 P 使得

$$PAP^{\mathrm{T}}=\begin{bmatrix}B & C\\ O & D\end{bmatrix},$$

则 $C=O$. 更一般地，A 有法式

$$QAQ^{\mathrm{T}}=\mathrm{diag}[A_{11},\cdots,A_{mm}],$$

其中诸 A_{jj} 为不可约双随机矩阵，Q 为某置换矩阵.

本定理的证明留给读者(见习题 1).

今后，用 Ω_n 表示所有 n 阶双随机矩阵的集合. 显然，Ω_n 也是 $\boldsymbol{M}_n(\mathbb{R})$ 中的凸体. 我们将证明属于 Birkhoff 的一个基本结果：凸体 Ω_n 以 $n!$ 个 n 阶置换矩阵为其顶点. 为此先证明关于矩阵的零-非零型的 Frobenius-König 定理.

设 $A=(a_{ij})$ 为 n 阶矩阵，且 σ 为集合 $\mathcal{N}=\{1,\cdots,n\}$ 的某个置换. 则 $(a_{1\sigma(1)},a_{2\sigma(2)},\cdots,a_{n\sigma(n)})$ 称为 A 的(对应于置换 σ 的)**对角线**. 回忆一下，如 k 与 n 为正整数且 $k\leqslant n$，则 $Q_{k,n}$ 表示递增序列 $\omega=(\omega_1,\cdots,\omega_k)$，$1\leqslant\omega_1<\omega_2<\cdots<\omega_k\leqslant n$ 的集合，它计有 $\binom{n}{k}$ 个元素. 若 $\alpha\in Q_{h,m}$ 与 $\beta\in Q_{k,n}$，A 为 $m\times n$ 矩阵，则 $A[\alpha|\beta]$ 表示 A 的 $h\times k$ 子阵，其 (i,j) 元素恰为 $a_{\alpha_i\beta_j}$. 类似地可定义 $A(\alpha|\beta)$，$A[\alpha|\beta)$ 与 $A(\alpha|\beta]$.

定理 2(Frobenius-König)　设 $A\in\boldsymbol{M}_n(\mathbb{R})$. 则 A 的每一条对角线至少含有一个零元素当且仅当 A 有一个 $s\times t$ 零子阵，$s+t=n+1$.

证明　设 A 的子阵 $A[\alpha|\beta]=O$，其中，$\alpha\in Q_{s,n}$，$\beta\in Q_{t,n}$，$s+t=n+1$. 若有 A 的某对角线不含零元素，则此对角线在列 β 中的元素必定位于 $A(\alpha|\beta]$ 之中. 但是，

$A(\alpha|\beta]$为$(n-s)\times t$的矩阵，因而$t\leqslant n-s$. 此与$t=n-s+1$矛盾. 反之，假设A的每一条对角线都至少含有一个零元素，要证A有$s\times t$零子阵，$s+t=n+1$. 对阶数n应用归纳法. 对$n=1$结论显然成立. 假定对小于n的情形结论成立，考虑n阶的情形. 当$A=O$时结果显然成立，因而可设A的某元素$a_{ij}\neq 0$. 这时$A(i|j)$的每一条对角线含有零元素，按归纳假定，$A(i|j)$有一个$u\times v$零子阵使得$u+v=n$. 于是，存在置换矩阵P与Q使得

$$PAQ=\begin{bmatrix} X & O \\ Y & Z \end{bmatrix}, \tag{1}$$

式中，X与Z分别为u阶与$n-u$阶矩阵. 显然，PAQ的任一对角线含有零元素. 若X(或Z)的某一对角线元素均非零，按题设则Z(或X)的所有对角线一定含有零元素，也就是说，X与Z中至少有一个矩阵其每一条对角线含有零元素. 不妨设矩阵X具有这个性质. 此时由归纳假定知道，X有一个$p\times q$零子阵，$p+q=u+1$. 于是PAQ这时有一个$p\times(q+v)$零子阵，$p+(q+v)=(p+q)+n-u=n+1$. □

应用上述定理，我们可以证明关于Ω_n结构的基本结果(在第六章6.3.2小节中曾用过这个结果).

定理3(Birkhoff)　n阶双随机矩阵的集合Ω_n是以$n!$个n阶置换矩阵为顶点的凸体.

证明　设$A\in\Omega_n$. 首先证明:A为n阶置换矩阵的凸组合:

$$A=\sum_{j=1}^{n!}\theta_j P_j,\quad \theta_j\geqslant 0,\quad \sum_{j=1}^{n!}\theta_j=1, \tag{2}$$

其中，各P_j为n阶置换矩阵. 为此，对A的正元素个数应用归纳法. 当A恰有n个正元素时，A本身为置换矩阵，因而(2)式成立. 现假定A不是n阶置换矩阵(因而A的正元素个数大于n)且有一个$s\times t$零矩阵$A[\omega|\tau]$，$\omega\in Q_{s,n}$与$\tau\in Q_{t,n}$. 由于A在行ω中的非零元素必在$A[\omega|\tau)$内，故$A[\omega|\tau)$每一行元素之和皆等于1，因而$A[\omega|\tau)$中所有元素之和等于s. 同理可证，$A(\omega|\tau]$中所有元素之和等于t. 但$A[\omega|\tau)$与$A(\omega|\tau]$为元素不重合的A的子矩阵，且因A中所有元素之和为n，所以有$s+t\leqslant n$. 这表明:A的任意的$s\times t$零矩阵不满足$s+t=n+1$. 依照定理2，A至少有一条对角线其元素均为正数. 现设P为某n阶置换矩阵，它的n个元素1对应这些正元素组成的对角线，且用θ表示这条对角线中的最小正数. 显然，$0<\theta<1$，且$T\equiv(A-\theta P)(1-\theta)^{-1}\in\Omega_n$，这里$T$比$A$至少少一个正元素. 按归纳假定，$T$为$n$阶置换矩阵的凸组合，因而

$$A=\theta P+(1-\theta)T$$

也为n阶置换矩阵的凸组合.

其次，每个n阶置换矩阵为Ω_n的顶点的事实是明显的，因为任一n阶置换矩阵不会是其他n阶置换矩阵的凸组合. □

由前面知道，若 $A\in\Omega_n$，则 $\rho(A)=1$ 为 A 的特征值. 对于 $A\in\Omega_n$ 的其他特征值，下列结果给出一种定位定理.

定理 4　设 $A=(a_{ij})\in\Omega_n$ 且 $1\neq\lambda\in\sigma(A)$. 则有

$$|\lambda|\leqslant\min\{1-\sum_j\min_i a_{ij},(\sum_j\max_i a_{ij})-1\}. \tag{3}$$

证明　设 $A^{\mathrm{T}}\boldsymbol{y}=\lambda\boldsymbol{y},\boldsymbol{y}\neq\boldsymbol{0}$. 从 $A\boldsymbol{e}=\boldsymbol{e}$ 推出，$\boldsymbol{y}^{\mathrm{T}}\boldsymbol{e}=\boldsymbol{y}^{\mathrm{T}}A\boldsymbol{e}=\lambda\boldsymbol{y}^{\mathrm{T}}\boldsymbol{e}$，因而 $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{e}=0$. 现令 $c_j=\min\limits_i a_{ij}$ 与 $A_c=(a_{ij}-c_j)\in\boldsymbol{M}_n(\mathbb{R})$. 则有 $A_c\geqslant O$ 与 $\lambda\boldsymbol{y}^{\mathrm{T}}=\boldsymbol{y}^{\mathrm{T}}A=\boldsymbol{y}^{\mathrm{T}}A_c$，于是，

$$|\lambda||\boldsymbol{y}^{\mathrm{T}}|\leqslant|\boldsymbol{y}^{\mathrm{T}}|A_c,\ |\lambda|\leqslant\rho(A_c).$$

但 A_c 的所有行和都等于 $1-\sum\limits_{j=1}^n c_j$，按 7.2.1 小节定理 6，$\rho(A_c)=1-\sum\limits_{j=1}^n c_j=1-\sum\limits_{j=1}^n\min\limits_i a_{ij}$. 同理，若令 $d_j=\max\limits_i a_{ij}$ 与 $A_d=(d_j-a_{ij})\in M_n(\mathbb{R})$，则有 $|\lambda|\leqslant\rho(A_d)=\sum\limits_{j=1}^n d_j-1=(\sum\limits_{j=1}^n\max\limits_i a_{ij})-1$. 因此(3) 式成立. □

现在讨论关于双随机矩阵的另一个基本结果. 它主要表明：$\mathbb{R}^n$ 中向量 $\boldsymbol{z}$ **控制** $\boldsymbol{y}$ 当且仅当有 $\boldsymbol{D}\in\Omega_n$ 使得 $\boldsymbol{y}=D\boldsymbol{z}$. 这里，$\boldsymbol{z}\in\mathbb{R}^n$ 控制 $\boldsymbol{y}\in\mathbb{R}^n$（记为 $\boldsymbol{y}\prec\boldsymbol{z}$ 或 $\boldsymbol{z}\succ\boldsymbol{y}$），意指

$$\begin{aligned}&\max_{\omega\in Q_{k,n}}\sum_{j=1}^k y_{\omega(j)}\leqslant\max_{\omega\in Q_{k,n}}\sum_{j=1}^k z_{\omega(j)},\quad k=1,2,\cdots,n,\\&y_1+y_2+\cdots+y_n=z_1+z_2+\cdots+z_n.\end{aligned} \tag{4}$$

定理 5　设 $\boldsymbol{y}=(y_1,\cdots,y_n)^{\mathrm{T}},\boldsymbol{z}=(z_1,\cdots,z_n)^{\mathrm{T}}\in\mathbb{R}^n$. 则下列性质彼此等价：

(1) $\boldsymbol{y}\prec\boldsymbol{z}$.

(2) 存在 n 阶实对称矩阵，它以 $z_1,\cdots,z_n$ 为特征值，以 $y_1,\cdots,y_n$ 为主对角元素.

(3) 存在 n 阶 Hermite 矩阵，它以 $z_1,\cdots,z_n$ 为特征值，以 $y_1,\cdots,y_n$ 为主对角元素.

(4) 存在 n 阶实正交矩阵 $C=(c_{ij})$ 使得 $\boldsymbol{y}=G\boldsymbol{z}$，这里 $G=(c_{ij}^2)\in\boldsymbol{M}_n(\mathbb{R})$.

(5) 存在 n 阶酉矩阵 $U=(u_{ij})$ 使得 $\boldsymbol{y}=H\boldsymbol{z}$. 这里 $H=(|u_{ij}|^2)\in\boldsymbol{M}_n(\mathbb{R})$.

(6) 存在 $D\in\Omega_n$ 使得 $\boldsymbol{y}=D\boldsymbol{z}$.

证明　我们将按下列蕴涵关系证明此六个性质的等价性：

$$(1)\Rightarrow(2)\Rightarrow(3)\Rightarrow(5)\Rightarrow(6)\Rightarrow(1);(2)\Rightarrow(4)\Rightarrow(5).$$

不失普遍性，假定 $y_1\geqslant y_2\geqslant\cdots\geqslant y_n$ 与 $z_1\geqslant z_2\geqslant\cdots\geqslant z_n$.

(1)⇒(2)：对维数 n 应用归纳法. 对 $n=1$ 结果显然成立. 对 $n=2$，$\boldsymbol{y}\prec\boldsymbol{z}$ 等价于 $y_1\leqslant z_1$ 与 $y_1+y_2=z_1+z_2$. 这时，直接验证知道，二阶实对称矩阵

$$\begin{bmatrix}y_1&w\\w&y_2\end{bmatrix}$$

(式中,$w=\sqrt{(z_1-y_1)(z_1-y_2)}$)有特征值 z_1 与 z_2. 现令 $n>2$,并假定(1)⇒(2)对任意 $2m$ 个实数成立,$1\leqslant m<n$. 记

$$p=\min_{1\leqslant s\leqslant n-1}\left(\sum_{i=1}^{s}z_i-\sum_{i=1}^{s}y_i\right), \tag{5}$$

并假定当 $s=t$ 时此最小值达到,$1\leqslant t\leqslant n-1$. 由于 $z_1-y_1\geqslant p$ 与 $\sum_{i=1}^{n-1}z_i-\sum_{i=1}^{n-1}y_i=-z_n+y_n\geqslant p$,故有 $z_1-z_n\geqslant 2p+y_1-y_n$,因而 $z_1-z_n\geqslant 2p$. 令 $z'_1=z_1-p$, $z'_j=z_j(2\leqslant j\leqslant n-1)$, $z'_n=z_n+p$. 则有

$$\begin{aligned}&\sum_{i=1}^{s}y_i\leqslant\sum_{i=1}^{s}z'_i,\qquad s=1,\cdots,t-1,\\&\sum_{i=1}^{t}y_i=\sum_{i=1}^{t}z'_i,\end{aligned} \tag{6}$$

因为 $p=\sum_{i=1}^{t}z_i-\sum_{i=1}^{t}y_i$,所以同时有,

$$\begin{aligned}&\sum_{i=t+1}^{s}y_i\leqslant\sum_{i=t+1}^{s}z'_i,\qquad s=t+1,\cdots,n-1\\&\sum_{i=t+1}^{n}y_i=\sum_{i=t+1}^{n}z'_i.\end{aligned} \tag{7}$$

(6)式蕴涵 $(y_1,\cdots,y_t)^{\mathrm{T}}$ 与 $(z'_1,\cdots,z'_t)^{\mathrm{T}}$ 满足(1)的性质,尽管 $z'_1\geqslant z'_2\geqslant\cdots\geqslant z'_t$ 可能不成立.(这是因为若 $z'_{q_1}\geqslant z'_{q_2}\geqslant\cdots\geqslant z'_{q_t}$,则 $y_1\leqslant z'_1\leqslant z'_{q_1}$, $y_1+y_2\leqslant z'_1+z'_2\leqslant z'_{q_1}+z'_{q_2}$, $\cdots$, $y_1+y_2+\cdots+y_t=z'_{q_1}+z'_{q_2}+\cdots+z'_{q_t}$)按归纳假定,存在 t 阶实对称矩阵 A_1,以 $z'_1,\cdots,z'_t$ 为特征值,以 $y_1,\cdots,y_t$ 为其主对角元素. 类似地,由(7)式推出,存在 $n-t$ 阶实对称矩阵 A_2,以 $z'_{t+1},\cdots,z'_n$ 为特征值,以 $y_{t+1},\cdots,y_n$ 为主对角元素. 由于 $z'_1=z_1-p\geqslant z_n+p=z'_n$, $z'_1\leqslant z_1$ 与 $z'_1+z'_n=z_1+z_n$,故按对 $n=2$ 已证过的命题知道,存在实数 w,使得二阶矩阵

$$\begin{bmatrix}z'_1 & w\\ w & z'_n\end{bmatrix}$$

有特征值 z_1 与 z_n. 现用 $\boldsymbol{u}\in\mathbb{R}^t$ 与 $\boldsymbol{v}\in\mathbb{R}^{n-t}$ 分别表示 A_1 对应于特征值 z'_1 的单位欧氏长度的特征向量与 A_2 对应于特征值 z'_n 的单位欧氏长度的特征向量. 根据习题 2,n 阶实对称矩阵

$$\begin{bmatrix}A_1 & w\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\\ w\boldsymbol{v}\boldsymbol{u}^{\mathrm{T}} & A_2\end{bmatrix} \tag{8}$$

有特征值 $z_1,z_2,\cdots,z_n$ 与主对角元素 $y_1,y_2,\cdots,y_n$.

(2)⇒(3):这是显然的.

(3)⇒(5):设 n 阶 Hermite 矩阵 $A=(a_{ij})$ 满足(3). 则有 $U=(u_{ij})\in\mathscr{U}_n$ 使得

$A=U\mathrm{diag}[z_1,\cdots,z_n]U^*$. 由于 $a_{ii}=y_i$，$1\leqslant i\leqslant n$，我们有 $y_i=a_{ii}=\sum_{j=1}^{n}|u_{ij}|^2 z_j$，$1\leqslant i\leqslant n$，此等价于 $\boldsymbol{y}=H\boldsymbol{z}$，$H=(|u_{ij}|^2)$.

(5)⇒(6)：此由 $H\in\Omega_n$ 事实立即得出.

(6)⇒(1)：设 $\boldsymbol{y}=D\boldsymbol{z}$，$D\in\Omega_n$. 此时，首先有

$$\sum_{i=1}^{n}y_i=\sum_{i=1}^{n}\sum_{j=1}^{n}d_{ij}z_j=\sum_{j=1}^{n}\left(\sum_{i=1}^{n}d_{ij}\right)z_j=\sum_{j=1}^{n}z_j.$$

其次，对任一固定 s，$1\leqslant s\leqslant n-1$，令

$$v_j=\sum_{i=1}^{s}d_{ij},\qquad j=1,\cdots,n.$$

显然，$0\leqslant v_j\leqslant 1$ 与 $\sum_{j=1}^{n}v_j=s$. 由于 $z_1\geqslant z_2\geqslant\cdots\geqslant z_n$，故有

$$\begin{aligned}\sum_{i=1}^{s}y_i&=\sum_{i=1}^{s}\sum_{j=1}^{n}d_{ij}z_j=\sum_{j=1}^{n}v_jz_j\leqslant\sum_{j=1}^{s-1}v_jz_j+\left(\sum_{j=s}^{n}v_j\right)z_s\\&=\sum_{j=1}^{s-1}v_jz_j+\left(s-\sum_{j=1}^{s-1}v_j\right)z_s=\sum_{j=1}^{s-1}v_j(z_j-z_s)+sz_s\\&\leqslant\sum_{j=1}^{s-1}(z_j-z_s)+sz_s=\sum_{i=1}^{s}z_i.\end{aligned}$$

(2)⇒(4)：设 n 阶实对称矩阵 $A=(a_{ij})$ 满足(2). 这时存在实正交矩阵 $C=(c_{ij})$ 使得 $A=C\mathrm{diag}[z_1,\cdots,z_n]C^{\mathrm{T}}$. 于是，

$$y_i=a_{ii}=\sum_{j=1}^{n}c_{ij}^2z_j,\qquad i=1,\cdots,n,$$

此即 $\boldsymbol{y}=G\boldsymbol{z}$，其中 $G=(c_{ij}^2)\in\Omega_n$.

(4)⇒(5)：这是显然的. □

强调一下从上述定理中(1)与(2)(或者(1)与(3))的等价性，便能直接地得到下列推论，它给出 Hermite(或实对称)矩阵的特征值与主对角元素之间的控制不等式关系.

推论 6　若 $A=(a_{ij})$ 为 n 阶 Hermite 或实对称矩阵，$a_{11}\geqslant a_{22}\geqslant\cdots\geqslant a_{nn}$，且 $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_n$ 为 A 的特征值，则 $(\lambda_1,\lambda_2,\cdots,\lambda_n)^{\mathrm{T}}\succ(a_{11},a_{22},\cdots,a_{nn})^{\mathrm{T}}$. 反之，若 $a_{11}\geqslant a_{22}\geqslant\cdots\geqslant a_{nn}$，$\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_n$，并且，$(\lambda_1,\lambda_2,\cdots,\lambda_n)^{\mathrm{T}}\succ(a_{11},a_{22},\cdots,a_{nn})^{\mathrm{T}}$，则存在 Hermite 或实对称矩阵以 $a_{11},\cdots,a_{nn}$ 为其主对角元素，以 $\lambda_1,\cdots,\lambda_n$ 为其特征值.

我们指出，定理 5 中(1)与(6)的等价性表明，若 $\boldsymbol{y}$ 与 $\boldsymbol{z}$ 满足 $\boldsymbol{y}=D\boldsymbol{z}$，$D\in\Omega_n$，则 $\boldsymbol{y}$ 的最大元素不超过 $\boldsymbol{z}$ 的最大元素，而 $\boldsymbol{y}$ 的最小元素不小于 $\boldsymbol{z}$ 的最小元素. 类似的事实对两个最大元素与两个最小元素之和也成立，余类推. 事实上，若 $y_1\geqslant y_2\geqslant\cdots\geqslant y_n$ 与 $z_1\geqslant z_2\geqslant\cdots\geqslant z_n$，则 $\boldsymbol{y}\prec\boldsymbol{z}$ 意味着 $y_1\leqslant z_1$，$y_1+y_2\leqslant z_1+z_2$，$\cdots$，$y_1+y_2+\cdots+y_{n-1}\leqslant z_1+z_2+\cdots+z_{n-1}$，$y_1+y_2+\cdots+y_n=z_1+z_2+\cdots+z_n$. 将这些不等式

改写为等价形式：

$$\begin{gathered} y_n \geqslant z_n, \\ y_n + y_{n-1} \geqslant z_n + z_{n-1}, \\ \vdots \\ y_n + y_{n-1} + \cdots + y_2 \geqslant z_n + z_{n-1} + \cdots + z_2, \\ y_n + y_{n-1} + \cdots + y_1 = z_n + z_{n-1} + \cdots + z_1. \end{gathered}$$

因此,一个双随机矩阵 $D\in\Omega_n$ 左乘$\mathbb{R}^n$ 中向量 $\boldsymbol{z}$ 可以起到"均衡"该向量诸元素大小的作用.

最后,顺便地介绍一下与双随机矩阵有关的著名的 **van der Waerden 猜想**. 1926 年 van der Waerden 提出如下猜想：

$$\min_{A\in\Omega_n} \operatorname{per} A = n!/n^n, \tag{9}$$

且只在所有元素等于 $1/n$ 的 n 阶双随机矩阵处达到最小值. 这里,对 n 阶矩阵 $A=(a_{ij})$, $\operatorname{per}A$ 表示 A 的**积和式**(permanent),亦即

$$\operatorname{per} A = \sum_{\sigma\in S_n} \prod_{i=1}^{n} a_{i\sigma(i)}, \tag{10}$$

式中,S_n 表示 n 阶对称群. 显然,$\max\limits_{A\in\Omega_n} \operatorname{per}A=1$,且最大值在每个 n 阶置换矩阵处达到. 经过半个世纪众多数学家的努力,此猜想终于在 1980 年为前苏联数学家 G. P. Egoritsjev 所证实. 对此有兴趣的读者可参阅文献[7],[18]与[13].

习题 7.3.2

1. 证明定理 1.

2. 设 A 与 B 分别为 m 阶与 n 阶 Hermite 矩阵或实对称矩阵,特征值分别为 $\alpha_1,\cdots,\alpha_m$ 与 $\beta_1,\cdots,\beta_n$,又设 $A\boldsymbol{u}=\alpha_1\boldsymbol{u}$, $\|\boldsymbol{u}\|_2=1$ 与 $B\boldsymbol{v}=\beta_1\boldsymbol{v}$, $\|\boldsymbol{v}\|_2=1$. 试证:对任意复数 η, $m+n$ 阶矩阵

$$C=\begin{bmatrix} A & \eta\boldsymbol{u}\boldsymbol{v}^* \\ \bar{\eta}\boldsymbol{v}\boldsymbol{u}^* & B \end{bmatrix}$$

有特征值 $\alpha_2,\cdots,\alpha_m,\beta_2,\cdots,\beta_n,\nu_1,\nu_2$,其中 ν_1 与 ν_2 为下列二阶矩阵的特征值:

$$\hat{C}=\begin{bmatrix} \alpha_1 & \eta \\ \bar{\eta} & \beta_1 \end{bmatrix}.$$

3. 试应用定理 3 与定理 5 证明:如果 $\boldsymbol{y},\boldsymbol{z}\in\mathbb{R}^n$ 满足(4)式中第一个不等式,且 $\varphi:\mathbb{R}^n\to\mathbb{R}$ 为对称度规函数,则 $\varphi(\boldsymbol{y})\leqslant\varphi(\boldsymbol{z})$. 这里,一个度规函数 $\varphi:\mathbb{R}^n\to\mathbb{R}$(见第二章习题 2.1.2 题 8(1))叫作**对称的**,假如对所有 $\boldsymbol{x}=(x_1,\cdots,x_n)^{\mathrm{T}}\in\mathbb{R}^n$, $\varphi(\boldsymbol{x})=\varphi(x_1,\cdots,x_n)=\varphi(\lambda_1x_1,\cdots,\lambda_nx_n)=\varphi(x_{\sigma(1)},\cdots,x_{\sigma(n)})$,式中诸 $\lambda_i=1$ 或 -1,而 σ 为 n 阶对称群 S_n 的任意置换.

4. 设 $A\in\Omega_n$. 试证:A 至多为 n^2-2n+2 个 n 阶置换矩阵的凸组合.

5. 试证:若 $A\in\Omega_n$,且 $D=\operatorname{diag}[d_1,d_2,\cdots,d_n]$为正对角矩阵,则 $\rho(DA)\geqslant(d_1d_2\cdots d_n)^{1/n}$.

6. 若 $A=(a_{ij})\in\mathcal{N}_n$的所有行元素之和都等于 s,则称 A 为**广义随机矩阵**. 试证:假如 A 的所

有对角元素中最小两个元素为 a_{kk} 与 a_{ll}，则 A 的所有特征值位于卵形

$$|z-a_{kk}||z-a_{ll}|\leqslant(s-a_{kk})(s-a_{ll})$$

的边界或内部.

7. 试证：$A=(a_{ij}):a_{ij}=1/n, \forall i,j\in\mathcal{N}$ 为唯一的 n 阶不可约幂等的双随机矩阵.

8. 设 $A\in\Omega_n$. 试证：$\mathrm{per}A>0$.

参考文献

[1] Berman A, Plemmons R J. Nonnegative Matrices in the Mathematical Sciences. New York: Academic Press, 1979, Reprinted and updated, Philadelphia: SIAM Press, 1994

[2] Rothblum U G. Algebraic eigenspaces of nonnegative matrices. Linear Algebra Appl., 1975(12), 281～292

[3] Rockafellar R T. Convex Analysis. Princeton. New Jersey: Princeton Univ. Press. 1970

[4] Pearl M. Matrix Theory and Finite Mathematics. New York: McGraw-Hill, 1973

[5] Horn A. Doubly stochastic matrices and the diagonal of a rotation matrix. Amer. J. Math., 1954(76), 620～630

[6] Mirsky L. Inequalities and existence theorems in the theory of matrices. J. Math. Anal. Appl., 1964(9), 99～118

[7] van Lint J H. Notes on Egoritsjev's proof of the van der Waerden conjecture. Linear Algebra Appl., 1981(39), 1～8

[8] Seneta E. Non-negative Matrices. New York: Wiley, 1973

[9] Gantmacher F R. Applications of the Theory of Matrices. New York: Interscience, 1959

[10] Brauer A. On the characteristic roots of non-negative matrices. see Hans Schneider. Recent Advances in Matrix Theory. Madison and Milwaukee, Univ. of Wisconsin Press, 1964, 3～38

[11] Cooper D H. On the maximum eigenvalue of a reducible nonnegative real matrix. Math. Z., 1973(131), 213～217

[12] Marshall A W, Olkin I. Inequalities: Theory of Majorization and Its Applications. New York: Academic Press, 1979

[13] Ando T. Majorization, Doubly Stochastic Matrices and Comparison of Eigenvalues. Lecture Note, Sapporo, Japan, 1982

[14] Wielandt H. Unzerlegbare, nicht-negative Matrizen. Math. Z., 1950(52), 642～648

[15] Seneta E. Non-negative Matrices and Markov Chains. New York: Springer-Verlag, 1981

[16] Minc H. Nonnegative Matrices. New York: Wiley, 1988

[17] Bapat R B, Raghavan T E S. Nonnegative Matrices and Applications. Cambridge: Cambridge University Press, MA, 1997

[18] Egorycev G B. A Solution of van der Waerden's permanent problem. Dok, Akad. Nauk SSSR, 1981(258), 1041～1044

第八章　*M*-矩阵

在计算数学、生物学、物理学与数理经济学等领域中，有许多问题可以归结为具有特殊构造的矩阵问题. 在这一类矩阵之中，具有非正的非对角元素的实方阵扮演着重要角色. 这些矩阵有如下形式：

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \in \boldsymbol{M}_n(\mathbb{R}),$$

式中，$a_{ij} \leqslant 0, \forall i \neq j$. 上述矩阵 A 可表示为

$$A = sI - B, \qquad s > 0 \text{ 与 } B \geqslant 0.$$

事实上，若所有 $a_{ii}=0$，可取 $s=1$，否则取 $s=\max\limits_i |a_{ii}|$，则有 $B=(b_{ij})$：$b_{ii}=s-a_{ii}\geqslant 0$，$b_{ij}=-a_{ij}\geqslant 0, \forall i\neq j$. 因此，这毫不奇怪地预示着非负矩阵在这些矩阵研究中的基本作用. 按传统习惯，我们引入记号

$$\mathbb{Z}^{n\times n} = \{A = (a_{ij}) \in \boldsymbol{M}_n(\mathbb{R}) : a_{ij} \leqslant 0, \forall i \neq j\}.$$

在本章主要研究$\mathbb{Z}^{n\times n}$的一个子集——*M*-矩阵集合的性质及其应用以及与它有关联的$\boldsymbol{M}_n(\mathbb{R})$的几个重要子集的特征.

8.1　非奇异 *M*-矩阵

M-矩阵的术语在 1937 年首先由 Ostrowski 提出，此术语关系到 Minkowski 在 1900 年与 1907 年的研究工作，这些工作证明：若 $A\in \mathbb{Z}^{n\times n}$ 的所有行元素之和为正数，则 $\det A>0$. 随后主要有两部分人大大地发展了 Ostrowski 的早期工作. 一是数学家，另一是经济学家. 数学家重视 *M*-矩阵在非负矩阵特征值界的估计与高阶稀疏线性方程组迭代解法的收敛性的估计中的应用，而经济学家主要从全替代性(gross substutability)，一般均衡的稳定性(stability of a general equilibrium)以及经济系统的 Leontief 投入-产出分析诸方面研究 *M*-矩阵理论(见本章 8.3 节).

在本节主要介绍非奇异 *M*-矩阵的定义与它的众多的等价性条件. 奇异 *M*-矩阵将放在下一节讨论.

定义　设 $A\in \boldsymbol{M}_n(\mathbb{R})$可以表示为

$$A = sI - B, \qquad s > 0, B \geqslant 0. \tag{$*$}$$

若 $s\geqslant\rho(B)$，则称 A 为 ***M*-矩阵**；若 $s>\rho(B)$，则称 A 为**非奇异 *M*-矩阵**.

显然，形如（ * ）式的矩阵属于集合$\mathbb{Z}^{n\times n}$.

8.1.1 主子式皆为正实数的实方阵

我们用$\boldsymbol{P}$表示$\boldsymbol{M}_n(\mathbb{R})$中主子式皆为正实数的矩阵集合. 显然，该集合包含$\boldsymbol{M}_n(\mathbb{R})$中的所有对称正定矩阵. 在后面，我们将证明：$A\in\boldsymbol{P}\cap\boldsymbol{Z}^{n\times n}$当且仅当$A$为非奇异$M$-矩阵(见8.1.2小节推论4).

先讨论两组只与矩阵主子式性质有关的条件的等价性.

定理 1 设$A=(a_{ij})\in\boldsymbol{M}_n(\mathbb{R})$. 则下列两命题彼此等价：

(1) A的所有前主子式为正数.

(2) 存在下三角形矩阵L与上三角形矩阵U，它们的对角元素都为正数，且有$A=LU$.

这是著名的**矩阵三角分解定理**(第一章1.1.1小节定理5)的一个直接推论.

命题 2 若$A=(a_{ij})\in\boldsymbol{Z}^{n\times n}$有三角分解$A=LU$，这里$L$与$U$分别为$n$阶下与上三角形矩阵，其对角元素皆为正数，则$L,U\in\boldsymbol{Z}^{n\times n}$.

证明 令$L=(l_{ij})$与$U=(u_{ij})$. 于是，$l_{ij}=0,\forall i<j$与$u_{ij}=0,\forall i>j$，且$l_{ii}>0$与$u_{ii}>0,\forall i\in\mathcal{N}$. 我们要证明：$l_{ij}\leqslant0$与$u_{ij}\leqslant0,\forall i\neq j$. 为此对$i+j$应用数学归纳法. 对$i+j=3$，$l_{21}\leqslant0$与$u_{12}\leqslant0$由$a_{12}=l_{11}u_{12}$与$a_{21}=l_{21}u_{11}$直接推得. 现假定$i+j\geqslant4(i\neq j)$且$l_{ks}\leqslant0$与$u_{ks}\leqslant0,\forall k+s<i+j(k\neq s)$. 此时，若$i<j$，则在

$$a_{ij}=l_{ii}u_{ij}+\sum_{k<i}l_{ik}u_{kj}\leqslant0$$

中，由于$l_{ik}\leqslant0$与$u_{kj}\leqslant0,\forall k<i$，我们有$\sum\limits_{k<i}l_{ik}u_{kj}\geqslant0$，因而有$u_{ij}\leqslant0,\forall i<j$. 类似地，若$i>j$，则从关系式

$$a_{ij}=u_{jj}l_{ij}+\sum_{k<j}l_{ik}u_{kj}\leqslant0$$

也能得出$l_{ij}\leqslant0,\forall i>j$. □

定理 3 设$A=(a_{ij})\in\boldsymbol{M}_n(\mathbb{R})$. 则下列诸条件是彼此等价的：

(1) $A\in\boldsymbol{P}$.

(2) 若$D\in\boldsymbol{M}_n(\mathbb{R})$为非负对角矩阵，则$A+D$的所有主子式不为零.

(3) 对任意$\boldsymbol{0}\neq\boldsymbol{x}\in\mathbb{R}^n$，且$\boldsymbol{y}=A\boldsymbol{x}=(y_1,\cdots,y_n)^{\mathrm{T}}$，存在$i\in\mathcal{N}$使得$x_iy_i>0$(即$A$不颠倒任意向量的符号).

(4) 对任意$\boldsymbol{0}\neq\boldsymbol{x}\in\mathbb{R}^n$，存在正对角矩阵$D_x$使得

$$(A\boldsymbol{x},D_x\boldsymbol{x})=\boldsymbol{x}^{\mathrm{T}}D_xA\boldsymbol{x}>0.$$

(5) 对任意$\boldsymbol{0}\neq\boldsymbol{x}\in\mathbb{R}^n$，存在非负对角矩阵$H_x$使得

$$(A\boldsymbol{x},H_x\boldsymbol{x})=\boldsymbol{x}^{\mathrm{T}}H_xA\boldsymbol{x}>0.$$

(6) A与A的每一个主子阵的任意实的特征值为正数.

(7) 对任意 $\mathbf{0}\neq\boldsymbol{x}\in\mathbb{R}^n$，且 $\boldsymbol{z}=A^{\mathrm{T}}\boldsymbol{x}=(z_1,\cdots,z_n)^{\mathrm{T}}$，存在 $k\in\mathcal{N}$ 使得 $x_kz_k>0$.

(8) 对任意 n 阶**符号矩阵** S，即 S 为对角元素是 1 或 -1 的对角矩阵，$SA^{\mathrm{T}}S\boldsymbol{z}\leqslant\mathbf{0}$ 与 $\boldsymbol{z}\geqslant\mathbf{0}$ 蕴涵 $\boldsymbol{z}=\mathbf{0}$.

(9) 对任意 n 阶符号矩阵 S，存在 $\boldsymbol{x}\geqslant\mathbf{0}$ 使得 $SAS\boldsymbol{x}>\mathbf{0}$.

证明　首先证明性质(1)～(6)的等价性.

(1)⇒(2)：由第一章 1.1.1 小节(17)式直接推出.

(2)⇒(3)：用反证法. 设有 $\boldsymbol{x}=(x_1,\cdots,x_n)^{\mathrm{T}}\neq\mathbf{0}$ 满足 $x_iy_i\leqslant0,\forall i\in\mathcal{N}$，其中 $\boldsymbol{y}=(y_1,\cdots,y_n)^{\mathrm{T}}=A\boldsymbol{x}$. 令 $\mathcal{M}=\{i\in\mathcal{N}:x_i\neq0\}=\{i_1,i_2\cdots,i_m\}$，其中 $1\leqslant i_1<i_2<\cdots<i_m\leqslant n,1\leqslant m\leqslant n$，且 $A[\mathcal{M}]$ 表示行与列取自 $\mathcal{M}$ 的 A 的主子阵，$\boldsymbol{x}(\mathcal{M})=(x_{i_1},\cdots,x_{i_m})^{\mathrm{T}}$ 与 $\boldsymbol{y}(\mathcal{M})=(y_{i_1},\cdots,y_{i_m})^{\mathrm{T}}$. 则有 $\boldsymbol{y}(\mathcal{M})=A[\mathcal{M}]\boldsymbol{x}(\mathcal{M})$，并且存在非负对角矩阵 $D[\mathcal{M}]$ 使得 $\boldsymbol{y}(\mathcal{M})=-D[\mathcal{M}]\boldsymbol{x}(\mathcal{M})$. 于是，$(A[\mathcal{M}]+D[\mathcal{M}])\boldsymbol{x}(\mathcal{M})=\mathbf{0}$ 蕴涵 $A[\mathcal{M}]+D[\mathcal{M}]$ 为奇异的，此与(2)矛盾.

(3)⇒(4)：设(3)成立，取 $\boldsymbol{x}=(x_1,\cdots,x_n)^{\mathrm{T}}\neq\mathbf{0}$ 且令 $\boldsymbol{y}=(y_1,\cdots,y_n)^{\mathrm{T}}=A\boldsymbol{x}$. 按(3)有 $i\in\mathcal{N}$ 使得 $x_iy_i>0$. 因此，存在充分小的正数 ε 使得 $x_iy_i+\varepsilon\sum\limits_{j\neq i}x_jy_j>0$. 令 $D_x=\mathrm{diag}[d_1,\cdots,d_n]:d_i=1;d_j=\varepsilon,\forall j\neq i$，则 $(A\boldsymbol{x},D_x\boldsymbol{x})>0$.

(4)⇒(5)：这是显然的.

(5)⇒(6)：设 $\mathcal{M}=\{i_1,\cdots,i_m\}$ 为 $\mathcal{N}$ 的子集，其中，$1\leqslant i_1<i_2<\cdots<i_m\leqslant n,1\leqslant m\leqslant n$，又设 $\lambda\in\sigma(A[\mathcal{M}])$，这里 $A[\mathcal{M}]$ 为对应于 $\mathcal{M}$ 的 A 的主子阵. 若 λ 为实数，则对应于 λ 的 $A[\mathcal{M}]$ 的特征向量 $\boldsymbol{x}(\mathcal{M})=(x_{i_1},\cdots,x_{i_m})^{\mathrm{T}}$ 可取为实向量. 令 $\tilde{\boldsymbol{x}}=(\tilde{x}_1,\cdots,\tilde{x}_n)^{\mathrm{T}}\in\mathbb{R}^n$ 使得 $\tilde{x}_k=x_k,\forall k\in\mathcal{M}$，且 $\tilde{x}_k=0,\forall k\notin\mathcal{M}$. 按(5)，有非负对角矩阵 $H=\mathrm{diag}[h_1,\cdots,h_n]$ 使得 $(A\tilde{\boldsymbol{x}},H\tilde{\boldsymbol{x}})>0$. 但如令 $H(\mathcal{M})=\mathrm{diag}[h_{i_1},\cdots,h_{i_m}]$，则 $(A\tilde{\boldsymbol{x}},H\tilde{\boldsymbol{x}})=(A[\mathcal{M}]\boldsymbol{x}(\mathcal{M}),H[\mathcal{M}]\boldsymbol{x}(\mathcal{M}))=\lambda(\boldsymbol{x}(\mathcal{M}),H[\mathcal{M}]\boldsymbol{x}(\mathcal{M}))=\lambda(\tilde{\boldsymbol{x}},H\tilde{\boldsymbol{x}})$. 因为 $(\tilde{\boldsymbol{x}},H\tilde{\boldsymbol{x}})\geqslant0$，我们有 $\lambda>0$.

(6)⇒(1)：因为 A 的主子式 $\det A[\mathcal{M}]$ 等于 $A[\mathcal{M}]$ 所有特征值的乘积，但实矩阵 $A[\mathcal{M}]$ 的所有非实特征值(如果存在的话)的乘积为正数，所以(1)成立.

(7)⇔(3)：由于(1)⇔(3)且它们等价于 $A^{\mathrm{T}}\in\boldsymbol{P}$，故有(7)等价于(3).

(7)⇔(8)：设(8)不成立，即有 n 阶符号矩阵 S 与 $\boldsymbol{u}\geqslant\mathbf{0}$ 但 $\boldsymbol{u}\neq\mathbf{0}$ 使得 $SA^{\mathrm{T}}S\boldsymbol{u}\leqslant\mathbf{0}$. 置 $\boldsymbol{y}=(y_1,\cdots,y_n)^{\mathrm{T}}=S\boldsymbol{u}$，于是 $\boldsymbol{y}\neq\mathbf{0}$. 这时，因为 $\boldsymbol{z}=A^{\mathrm{T}}\boldsymbol{y}=(z_1,\cdots,z_n)^{\mathrm{T}}$ 满足 $y_kz_k\leqslant0,\forall k\in\mathcal{N}$，所以(7)不成立. 反之，设(7)不成立，即有向量 $\boldsymbol{y}=(y_1,\cdots,y_n)^{\mathrm{T}}\neq\mathbf{0}$ 使得 $\boldsymbol{z}=A^{\mathrm{T}}\boldsymbol{y}=(z_1,\cdots,z_n)^{\mathrm{T}}$ 满足 $y_kz_k\leqslant0,\forall k\in\mathcal{N}$. 显然，存在 n 阶符号矩阵 S 使得 $\boldsymbol{u}=S\boldsymbol{y}\geqslant\mathbf{0}$ 与 $S\boldsymbol{z}\leqslant\mathbf{0}$. 因为 $SA^{\mathrm{T}}S\boldsymbol{u}\leqslant\mathbf{0}$ 与 $\boldsymbol{u}\neq\mathbf{0}$，所以(8)不成立.

(8)⇔(9)：这是下一个关于线性不等式的定理的直接推论. □

定理 4　设 $A=(a_{ij})\in\boldsymbol{M}_{m,n}(\mathbb{R})$. 则下列三个命题彼此等价：

(1) 存在 $\boldsymbol{x}\geqslant\mathbf{0}$ 使得 $A\boldsymbol{x}>\mathbf{0}$.

(2) 若 $\boldsymbol{y}\geqslant\boldsymbol{0}$ 与 $A^{\mathrm{T}}\boldsymbol{y}\leqslant\boldsymbol{0}$,则 $\boldsymbol{y}=\boldsymbol{0}$.

(3) 存在 $\boldsymbol{x}>\boldsymbol{0}$ 使得 $A\boldsymbol{x}>\boldsymbol{0}$.

证明　(1)⇒(2):设(1)成立且有 $\boldsymbol{y}\neq\boldsymbol{0}$ 满足 $\boldsymbol{y}\geqslant\boldsymbol{0}$ 与 $A^{\mathrm{T}}\boldsymbol{y}\leqslant\boldsymbol{0}$. 这时 $\boldsymbol{y}^{\mathrm{T}}A\boldsymbol{x}=\boldsymbol{y}^{\mathrm{T}}(A\boldsymbol{x})>0$,另一方面,$\boldsymbol{y}^{\mathrm{T}}A\boldsymbol{x}=(A^{\mathrm{T}}\boldsymbol{y})^{\mathrm{T}}\boldsymbol{x}\leqslant 0$,此推出矛盾.

(2)⇒(3):设(2)成立,并用 C_1 与 C_2 分别表示$\mathbb{R}_+^m$与$\mathbb{R}_+^n$(它们均为凸锥),$C_3=\{-A^{\mathrm{T}}\boldsymbol{x}:\boldsymbol{x}\in C_1\}$($C_3$ 还是一个凸锥). 按(2),$C_2\cap C_3=\{\boldsymbol{0}\}$. 因此,应用第二章 2.3.1 小节定理 4 及其后说明,在$\mathbb{R}^n$ 中有超平面 $\boldsymbol{u}^{\mathrm{T}}\boldsymbol{x}=0$ 使得所有 $\boldsymbol{0}\neq\boldsymbol{x}\in C_2$ 满足 $\boldsymbol{u}^{\mathrm{T}}\boldsymbol{x}>0$,所有 $\boldsymbol{0}\neq\boldsymbol{x}\in C_3$ 满足 $\boldsymbol{u}^{\mathrm{T}}\boldsymbol{x}<0$. 前一条件蕴涵 $\boldsymbol{u}>\boldsymbol{0}$,后一条件蕴涵对所有使得 $A^{\mathrm{T}}\boldsymbol{z}\neq\boldsymbol{0}$ 的向量 $\boldsymbol{z}\geqslant\boldsymbol{0}$,$\boldsymbol{u}^{\mathrm{T}}A^{\mathrm{T}}\boldsymbol{z}>0$. 由(2)对所有 $\boldsymbol{z}\geqslant\boldsymbol{0},\boldsymbol{z}\neq\boldsymbol{0}$,我们有 $A^{\mathrm{T}}\boldsymbol{z}\neq\boldsymbol{0}$,从而 $\boldsymbol{u}^{\mathrm{T}}A^{\mathrm{T}}\boldsymbol{z}>0$. 于是,$\boldsymbol{u}^{\mathrm{T}}A^{\mathrm{T}}>\boldsymbol{0}^{\mathrm{T}}$,即有 $A\boldsymbol{u}>\boldsymbol{0}$.

(3)⇒(1):这是显然的. □

假如在定理 3(4)中令 $D_x=I$,在(9)中令 $S=I$,我们可分别得到如下两个推论.

推论 5　设 $A\in\boldsymbol{M}_n(\mathbb{R})$. 若 $A+A^{\mathrm{T}}$ 为对称正定矩阵,则 $A\in\boldsymbol{P}$.

推论 6　若 $A\in\boldsymbol{P}$,则存在 $\boldsymbol{x}\geqslant\boldsymbol{0}$ 使得 $A\boldsymbol{x}>\boldsymbol{0}$.

此外,从 $\boldsymbol{P}$ 的定义容易看出如下结论成立.

推论 7　若 $A\in\boldsymbol{P}$,且 D 为 n 阶正对角矩阵,则 DA 与 AD 均属于 $\boldsymbol{P}$.

矩阵类 $\boldsymbol{P}$ 在所谓**线性互补问题**(linear complementarity problem)中有重要作用. 具体地说,给定 $M\in\boldsymbol{M}_n(\mathbb{R})$与 $\boldsymbol{q}\in\mathbb{R}^n$,该问题旨在寻求向量 $\boldsymbol{w}\geqslant\boldsymbol{0}$ 与 $\boldsymbol{z}\geqslant\boldsymbol{0}$ 满足

$$\begin{aligned}\boldsymbol{w}&=M\boldsymbol{z}+\boldsymbol{q},\\ \boldsymbol{w}^{\mathrm{T}}\boldsymbol{z}&=0.\end{aligned}\tag{1}$$

我们提出(但不给出证明)上述问题的存在性定理. 有关证明请参阅文献[1].

定理 8　设 $M\in\boldsymbol{M}_n(\mathbb{R})$. 则问题(1)对任意 $\boldsymbol{q}\in\mathbb{R}^n$ 有唯一解 $\boldsymbol{z}\geqslant\boldsymbol{0}$ 与 $\boldsymbol{w}\geqslant\boldsymbol{0}$ 当且仅当 $M\in\boldsymbol{P}$.

习题 8.1.1

1. 设 $A=(a_{ij})\in\boldsymbol{M}_{m,n}(\mathbb{R})$. 考虑下列四个情形:

(1) 存在 $\boldsymbol{x}>\boldsymbol{0}$ 使得 $A\boldsymbol{x}>\boldsymbol{0}$.

(2) 存在 $\boldsymbol{y}\geqslant\boldsymbol{0},\boldsymbol{y}\neq\boldsymbol{0}$,使得 $A^{\mathrm{T}}\boldsymbol{y}\leqslant\boldsymbol{0}$.

(3) 存在 $\boldsymbol{x}\geqslant\boldsymbol{0},\boldsymbol{x}\neq\boldsymbol{0}$,使得 $A\boldsymbol{x}\geqslant\boldsymbol{0}$.

(4) 存在 $\boldsymbol{y}>\boldsymbol{0}$ 使得 $A^{\mathrm{T}}\boldsymbol{y}<\boldsymbol{0}$.

试证:(1)与(2)两种情形有且仅有一种出现;(3)与(4)两种情形有且仅有一种出现.

2. 设 $A\in\boldsymbol{M}_n(\mathrm{R})$. 试证下列两个命题彼此等价:

(1) $A+\alpha I$ 非奇异,$\forall\alpha\geqslant 0$.

(2) A 的每一个实特征值为正数.

3. 试证:若 $A\in\boldsymbol{P}$,则 A 的所有 k 阶主子式之和为正数,$k=1,2,\cdots,n$. 并且,若 $A\in M_n(\mathrm{R})$

的所有 k 阶主子式之和为正数,$k=1,2,\cdots,n$,则题 2 中(1)或(2)成立.

8.1.2 非奇异 M-矩阵的若干特性

非奇异 M-矩阵有很多重要的特性. 在这里我们只讨论其中的一部分,余下部分中一些留做习题.

定理 1 设 $A\in Z^{n\times n}$为非奇异 M-矩阵,且 $D\in Z^{n\times n}$满足 $A\leqslant D$. 则

(1) A^{-1}与 D^{-1}存在,且 $A^{-1}\geqslant D^{-1}\geqslant O$.

(2) D 的每一个实特征值为正数.

(3) $\det D\geqslant \det A>0$.

证明 设 A 为非奇异 M-矩阵且有形式

$$A=sI-B,\quad s>\rho(B),\quad B\geqslant O. \tag{1}$$

对任意给定 $\omega\leqslant 0$,考虑矩阵 $C=A-\omega I=(s-\omega)I-B$. 这时,$s-\omega>\rho(B)$,因而$C\in Z^{n\times n}$也是非奇异 M-矩阵. 这个事实表明:非奇异 M-矩阵的每一个实特征值必定为正数. 由于 $D\in Z^{n\times n}$,故存在足够小的正数 σ 使得 $U\equiv I-\sigma D\geqslant O$. 此时,$V\equiv I-\sigma A\geqslant I-\sigma D=U\geqslant O$. 按第七章 7.2.1 小节定理 1,$V$ 的谱半径 $\rho(V)$ 为 V 的非负特征值,我们有

$$\det((1-\rho(V))I-\sigma A)=\det(V-\rho(V)I)=0,$$

因而$(1-\rho(V))/\sigma$ 为 A 的实特征值. 但由本证明前一部分结果知道,$1-\rho(V)>0$,即 $0\leqslant\rho(V)<1$. 应用 Neumann 引理,$(I-V)^{-1}=(\sigma A)^{-1}=I+V+V^2+\cdots$,于是$(\sigma A)^{-1}\geqslant O$ 因而 $A^{-1}\geqslant O$. 由于 $O\leqslant U^k\leqslant V^k$,$k=1,2,\cdots$,且 $\rho(U)\leqslant\rho(V)$,故有

$$O\leqslant(I-U)^{-1}=(\sigma D)^{-1}=I+U+U^2+\cdots\leqslant(\sigma A)^{-1},$$

于是,$A^{-1}\geqslant D^{-1}\geqslant O$,(1)得证. 现令 $\alpha\leqslant 0$,则 $D-\alpha I\geqslant A$. 按(1),$D-\alpha I$ 非奇异,因而 D 的所有实特征值为正数,(2)得证. 最后证明(3). 按前面分析,只须证明:若 $A\in Z^{n\times n}$的所有实特征值为正数,且 $D\in Z^{n\times n}$满足 $A\leqslant D$,则有(3)成立. 对矩阵阶数 n 应用归纳法. 对 $n=1$,(3)显然正确. 假定(3)对 k 阶矩阵情形成立,$1\leqslant k<n$,我们来考虑 n 阶矩阵的情形. 令 $A_1=A[(1,2,\cdots,n-1)]$与 $D_1=D[(1,2,\cdots,n-1)]$. 则 $A_1,D_1\in Z^{(n-1)\times(n-1)}$且 $A_1\leqslant D_1$. 由于矩阵

$$\widetilde{A}\equiv\begin{bmatrix}A_1 & \mathbf{0}\\ \mathbf{0}^{\mathrm{T}} & a_{nn}\end{bmatrix}\in Z^{n\times n}$$

满足 $A\leqslant\widetilde{A}$,按(2)$\widetilde{A}$ 的所有实特征值为正数,因而 A_1 的所有实特征值也是正数. 根据归纳假定,$\det D_1\geqslant\det A_1>0$. 由(1),$A^{-1}\geqslant D^{-1}\geqslant O$ 推得$(A^{-1})_{nn}\geqslant(D^{-1})_{nn}\geqslant 0$(这里$(A^{-1})_{nn}$表示 A^{-1}的(n,n)元素),或即

$$\det A_1/\det A\geqslant\det D_1/\det D\geqslant 0.$$

因此,$\det A>0$ 与 $\det D>0$,且有

$$\det D\geqslant\det D_1\det A/\det A_1\geqslant\det A>0. \qquad\square$$

注意，从定理 1 证明看出，若 $A,D\in\mathbb{Z}^{n\times n}$，且 $D\geqslant A$，则 A 为非奇异 M-矩阵蕴涵 D 也是非奇异 M-矩阵，且有 $\det D\geqslant\det A>0$. 这个结果在许多实际问题中十分有用. 此外，如定理 1 中假设的“非奇异 M-矩阵”改换为“$A\in\mathbb{Z}^{n\times n}$ 的每一个实特征值皆为正数”，则其结论(1)～(3)仍然成立. 因此，我们有

推论 2 设 $A\in\mathbb{Z}^{n\times n}$ 的每一个实特征值皆为正数，且 $D\in\mathbb{Z}^{n\times n}$ 满足 $A\leqslant D$. 则定理 1 中(1)～(3)仍然成立.

定理 3 设 $A\in\mathbb{Z}^{n\times n}$. 则下列诸条件彼此等价：

(1) 存在 $\boldsymbol{x}\geqslant\boldsymbol{0}$ 使得 $A\boldsymbol{x}>\boldsymbol{0}$.

(2) 存在 $\boldsymbol{x}>\boldsymbol{0}$ 使得 $A\boldsymbol{x}>\boldsymbol{0}$(半正性).

(3) 存在正对角矩阵 D 使得 AD 严格对角占优，且 AD 的所有对角元素为正数.

(4) 若 $B\in\mathbb{Z}^{n\times n}$ 且 $B\geqslant A$，则 B 非奇异.

(5) A 的任意主子阵的每一个实特征值为正数.

(6) A 的所有主子式为正数.

(7) 对每个 $k(1\leqslant k\leqslant n)$，$A$ 的所有 k 阶主子式之和为正数.

(8) A 的每一个实特征值为正数.

(9) A 为非奇异的 M-矩阵.

(10) 存在 A 的一种分裂 $A=P-Q$ 使得 $P^{-1}\geqslant O,Q\geqslant O$ 与 $\rho(P^{-1}Q)<1$.

(11) A 非奇异且 $A^{-1}\geqslant O$(逆正性).

(12) A 的所有前主子式为正数.

(13) 存在 n 阶下三角形矩阵 L 与上三角形矩阵 U，使得 L 与 U 的主对角元素皆为正数，且 $A=LU$.

(14) 存在下三角形矩阵 $L\in\mathbb{Z}^{n\times n}$ 与上三角形矩阵 $U\in\mathbb{Z}^{n\times n}$，使得 L 与 U 的主对角元素皆为正数，且 $A=LU$.

(15) 对 A 的任意分裂 $A=P-Q$，若 $P^{-1}\geqslant O$ 与 $Q\geqslant O$，则 $\rho(P^{-1}Q)<1$.

(16) 存在正对角矩阵 D 使得矩阵 $E=DAD^{-1}$ 的实部 $\mathrm{Re}E=\frac{1}{2}(E+E^{\mathrm{T}})$ 为对称正定矩阵.

(17) 存在正对角矩阵 H 使得矩阵 $C=AH$ 的实部 $\mathrm{Re}C=\frac{1}{2}(C+C^{\mathrm{T}})$ 为对称正定矩阵.

(18) A 的任意特征值的实部为正数(正稳定性).

证明 先证头十一个性质彼此等价.

(1)⇒(2)：设 $\boldsymbol{x}\geqslant\boldsymbol{0}$ 满足 $A\boldsymbol{x}>\boldsymbol{0}$. 令 $\boldsymbol{e}=(1,\cdots,1)^{\mathrm{T}}\in\mathbb{R}^n$. 由于 $A\boldsymbol{x}>\boldsymbol{0}$，故有 $\varepsilon>0$ 使得 $A\boldsymbol{x}+\varepsilon A\boldsymbol{e}>\boldsymbol{0}$. 这时，$\boldsymbol{x}+\varepsilon\boldsymbol{e}>\boldsymbol{0}$ 满足 $A(\boldsymbol{x}+\varepsilon\boldsymbol{e})>\boldsymbol{0}$.

(2)⇒(3)：设 $\boldsymbol{x}>\boldsymbol{0}$ 满足 $A\boldsymbol{x}>\boldsymbol{0}$. 令 $D=\mathrm{diag}[x_1,\cdots,x_n]$. 则有 $a_{ii}x_i>\sum\limits_{j\neq i}|a_{ij}|x_j, i=1,\cdots,n$, 因而 AD 严格对角占优且其所有对角元素为正数.

(3)⇒(8)：设 $D=\mathrm{diag}[d_1,\cdots,d_n]$ 为(3)中的矩阵，且 AD 严格对角占优，其所有对角元素为正数. 则 A 为广义严格对角占优的(见第六章 6.1.1 小节)且所有 $a_{ii}>0$. 因此，从第六章 6.1.1 小节定理 5 便知，A 的所有特征值的实部为正数，因而(8)成立.

(8)⇒(9)：设 $A=sI-B, s>0$ 与 $B\geqslant O$. 则 $s-\rho(B)$ 为 A 的实特征值，按(8)，它是正数，即 $s>\rho(B)$. 因此，A 为非奇异 M-矩阵.

(9)⇒(4)：见定理 1(1).

(4)⇒(5)：设 $\varnothing\neq\mathscr{M}\subset\mathscr{N}=\{1,2,\cdots,n\}$ 且 λ 为 $A[\mathscr{M}]$ 的实特征值. 我们要证：(4)蕴涵 $\lambda>0$. 假若 $\lambda\leqslant 0$，定义 $B=(b_{ij})\in\mathbb{Z}^{n\times n}$ 如下：

$$b_{ij}=\begin{cases}a_{ii}-\lambda, & \text{如 } i=j,\\ a_{ij}, & \text{如 } i,j\in\mathscr{M} \text{ 且 } i\neq j,\\ 0, & \text{其他情形.}\end{cases}$$

由于 $\lambda\leqslant 0$，我们有 $B\geqslant A$，因而按(4)，B 为非奇异的. 另一方面，$\det B=\prod\limits_{i\notin\mathscr{M}} b_{ii}\det B[\mathscr{M}]=0$，因为 $B[\mathscr{M}]=A[\mathscr{M}]-\lambda I[\mathscr{M}]$ 且 λ 为 $A[\mathscr{M}]$ 的特征值，此推出矛盾. 因此 $\lambda>0$.

(5)⇒(6)：这是显然的，因为实方阵的所有非实特征值的乘积为正数.

(6)⇒(7)：显然成立.

(7)⇒(8)：应用公式

$$\det(A-\lambda I)=(-\lambda)^n+c_1(-\lambda)^{n-1}+\cdots+c_n, \tag{2}$$

式中，c_k 为 A 的所有 k 阶主子式之和. 按(7)，所有 $c_k>0$，因而(2)式中的多项式不能有非正的实根，即 A 的所有实特征值为正数.

(9)⇒(10)：取 $P=sI$ 与 $Q=B$，这里 s 与 B 由(1)式确定. 这时，$P^{-1}\geqslant O, Q\geqslant O$ 与 $\rho(P^{-1}Q)=\rho(B/s)=\rho(B)/s<1$.

(10)⇒(11)：设(10)成立，则有 $A=P(I-C)$，其中 $C=P^{-1}Q$. 由于 $\rho(C)<1$，按 Neumann 引理，$A^{-1}=(I+C+C^2+\cdots)P^{-1}$，于是，从 $C=P^{-1}Q\geqslant O$ 推出 $A^{-1}\geqslant O$.

(11)⇒(1)：令 $\boldsymbol{x}=A^{-1}\boldsymbol{e}$，这里 $\boldsymbol{e}=(1,1,\cdots,1)^{\mathrm{T}}$. 按(11)，$A^{-1}\geqslant O$，因而 $\boldsymbol{x}\geqslant\boldsymbol{0}$，$A\boldsymbol{x}=\boldsymbol{e}>\boldsymbol{0}$.

(6)⇒(12)：显然.

(12)⇒(13)：见 8.1.1 定理 1.

(13)⇒(14)：见 8.1.1 命题 2.

(14)⇒(11)：按推论 2，我们有 $L^{-1}\geqslant O$ 与 $U^{-1}\geqslant O$，因而 $A^{-1}=U^{-1}L^{-1}\geqslant O$.

(11)⇒(15)：设 $A=P-Q, P^{-1}\geqslant O$ 与 $Q\geqslant O$. 则 $P^{-1}Q\geqslant O$. 按第七章 7.2.1

小节定理 1,存在 $z\geqslant\mathbf{0},z\neq\mathbf{0}$ 使得 $P^{-1}Qz=\sigma z,\sigma=\rho(P^{-1}Q)$,因而 $Qz=\sigma Pz$. 假若 $\sigma\geqslant 1$,则 $Az=(P-Q)z=(\sigma^{-1}-1)Qz\leqslant\mathbf{0}$. 但由于 $A^{-1}\geqslant O$,故有 $A^{-1}Az=z\leqslant\mathbf{0}$,因而 $z=\mathbf{0}$,推出矛盾. 因此,$\rho(P^{-1}Q)<1$.

(15)⇒(9):取正数 s 使得 $s\geqslant\max_i a_{ii}$,且令 $B=sI-A$. 则 $P=sI$ 与 $Q=B$ 满足(15)的假设条件,因而 $\rho(P^{-1}Q)=\rho(s^{-1}B)=\rho(B)/s<1$,即 $s>\rho(B)$. 因此 A 为非奇异 M-矩阵.

(9)⇒(16):设 $A=sI-B,B\geqslant 0$ 与 $s>\rho(B)$. 若 $B\in\mathcal{N}_n$ 不可约,则有 $\boldsymbol{u}=(u_1,\cdots,u_n)^{\mathrm{T}}>\mathbf{0}$ 使得 $B\boldsymbol{u}=\rho(B)\boldsymbol{u}$,且有 $\boldsymbol{v}=(v_1,\cdots,v_n)^{\mathrm{T}}>\mathbf{0}$ 使得 $B^{\mathrm{T}}\boldsymbol{v}=\rho(B)\boldsymbol{v}$. 现令 $D=\mathrm{diag}[d_1,\cdots,d_n]$ 与 $\boldsymbol{w}=(w_1\cdots,w_n)^{\mathrm{T}}>\mathbf{0}$,其中,$d_i=(v_i/u_i)^{1/2}$ 与 $w_i=(u_iv_i)^{1/2},i=1,\cdots,n$. 由于 $D^{-1}\boldsymbol{w}=\boldsymbol{u}$ 与 $D\boldsymbol{w}=\boldsymbol{v}$,故有

$$(DBD^{-1})\boldsymbol{w}=\rho(B)\boldsymbol{w},\qquad (DBD^{-1})^{\mathrm{T}}\boldsymbol{w}=\rho(B)\boldsymbol{w}.$$

令 $E=DAD^{-1}$. 则 $E\in\mathbb{Z}^{n\times n}$满足 $E\boldsymbol{w}=(s-\rho(B))\boldsymbol{w}$ 与 $E^{\mathrm{T}}\boldsymbol{w}=(s-\rho(B))\boldsymbol{w}$. 因此,我们有$\frac{1}{2}(E+E^{\mathrm{T}})\boldsymbol{w}=(s-\rho(B))\boldsymbol{w},\boldsymbol{w}>\mathbf{0}$. 从性质(2)与(6)的等价性知道,矩阵$\frac{1}{2}(E+E^{\mathrm{T}})$的所有主子式为正数,因而$\frac{1}{2}(E+E^{\mathrm{T}})$为对称正定的. 现假定 $B\in\mathcal{N}_n$ 可约,且有法式

$$PBP^{\mathrm{T}}=\begin{bmatrix}B_{11} & B_{12} & \cdots & B_{1m}\\ O & B_{22} & \cdots & B_{2m}\\ \vdots & & \ddots & \vdots\\ O & O & \cdots & B_{mm}\end{bmatrix},$$

式中 B_{ii} 或为一阶零矩阵或为不可约矩阵,$i=1,\cdots,m$,令 $G=PAP^{\mathrm{T}}$,则 $G=[G_{ik}]$ 为分块三角阵:$G_{ik}=O,\ \forall\, i>k,G_{ik}=-B_{ik},\ \forall\, i<k$,且 $G_{ii}=sI-B_{ii}$. 按前一部分关于不可约情形的证明,存在正对角矩阵 D_i 使得 $E_i=D_iG_{ii}D_i^{-1}$ 的实部$\frac{1}{2}(E_i+E_i^{\mathrm{T}})$为对称正定矩阵. 对 $\varepsilon>0$,定义分块对角矩阵 $L(\varepsilon)=\mathrm{diag}[D_i,\varepsilon^{-1}D_2,\cdots,\varepsilon^{1-m}D_m]$. 则

$$L(\varepsilon)GL(\varepsilon)^{-1}=\begin{bmatrix}E_1 & E_{12} & E_{13} & \cdots & E_{1m}\\ O & E_2 & E_{23} & \cdots & E_{2m}\\ \vdots & \vdots & \vdots & & \vdots\\ O & O & O & \cdots & E_m\end{bmatrix},\qquad (3)$$

式中 $E_{ik}=\varepsilon^{k-i}D_iG_{ik}D_k^{-1},\ \forall\, i<k$. 由于分块三角矩阵 $\mathrm{diag}[E_1,\cdots,E_m]$的实部为对称正定矩阵,故存在 $\varepsilon_0>0$ 使得当 $\varepsilon=\varepsilon_0$时,(3)式中矩阵的实部也为对称正定矩阵(见第一章习题 1.2.2 题 15). 记 $D=P^{\mathrm{T}}L(\varepsilon_0)P$. 这时,$E=DAD^{-1}$的实部也是对称正定矩阵,因为

$$E=P^{\mathrm{T}}L(\varepsilon_0)PAP^{\mathrm{T}}L(\varepsilon_0)^{-1}P=P^{\mathrm{T}}L(\varepsilon_0)GL(\varepsilon_0)^{-1}P,$$

因而 E 的实部为 $L(\varepsilon_0)GL(\varepsilon_0)^{-1}$ 实部行列进行同步变换后的结果.

(16)⇒(17)：设 D 为(16)中的正对角矩阵且 $H=D^{-2}$. 由于 DAD^{-1} 的实部为对称正定的，从

$$AH + HA^{\mathrm{T}} = D^{-1}(DAD^{-1} + D^{-1}A^{\mathrm{T}}D)D^{-1} \tag{4}$$

看出，AH 的实部也是对称正定矩阵.

(17)⇒(18)：设(17)成立，即 $\mathrm{Re}C=\frac{1}{2}(AH+HA^{\mathrm{T}})$ 为对称正定的. 这时，$((\mathrm{Re}C)\boldsymbol{x},\boldsymbol{x})>0$，$\forall\,\boldsymbol{x}\neq\boldsymbol{0}$，即有 $\mathrm{Re}(AH\boldsymbol{x},\boldsymbol{x})>0$，$\forall\,\boldsymbol{x}\neq\boldsymbol{0}$. 现设 $\lambda\in\sigma(A)$，$\boldsymbol{y}\neq\boldsymbol{0}$ 为对应的特征向量，且 $\boldsymbol{x}=H^{-1}\boldsymbol{y}$. 则 $\boldsymbol{x}\neq\boldsymbol{0}$，$\mathrm{Re}(A\boldsymbol{y},H^{-1}\boldsymbol{y})=\mathrm{Re}(\lambda\boldsymbol{y},H^{-1}\boldsymbol{y})>0$ 因而 $\mathrm{Re}\lambda>0$，因为 $(\boldsymbol{y},H^{-1}\boldsymbol{y})>0$.

(18)⇒(8)：显然. □

我们指出，上述定理为非奇异 M-矩阵的基本定理. 它提供众多的非奇异 M-矩阵的等价条件. 其中，许多条件又可以引伸出另一些等价条件. 通常，我们将满足该定理条件(2)的矩阵 $A\in\boldsymbol{M}_n(\mathbb{R})$（不必限定于 $\mathbb{Z}^{n\times n}$）叫作**半正矩阵**. 满足该定理条件(11)的矩阵 $A\in\boldsymbol{M}_n(\mathbb{R})$ 叫作**逆正的**（见第五章 5.2.3 小节定义 1）. $A\in\boldsymbol{M}_n(\mathbb{R})$ 为逆正的当且仅当 $A\in\boldsymbol{M}_n(\mathbb{R})$ 为单调的，即 $A\boldsymbol{x}\geqslant A\boldsymbol{y}$ 蕴涵 $\boldsymbol{x}\geqslant\boldsymbol{y}$（见第五章 5.2.3 小节定理 2）. 逆正或单调矩阵的其他等价条件见习题 4. 若 $A\in\boldsymbol{M}_n(\mathbb{R})$ 满足该定理条件(10)，则称 A 有**收敛的规则分裂** $A=P-Q$. 从习题 4 知道，当 $A\in\boldsymbol{M}_n(\mathbb{R})$ 时，A 有收敛的规则分裂当且仅当 A 为逆正的（或单调的）. 此外，当 $A\in\boldsymbol{M}_n(\mathbb{R})$ 时，该定理条件(18)等价于 $-A$ 为稳定矩阵（见第四章 4.2.1 小节定理 6）. 因此，通常将满足该定理条件(18)的 $A\in\boldsymbol{M}_n(\mathbb{R})$ 叫作**正稳定矩阵**. 最后，当 $A\in\boldsymbol{M}_n(\mathbb{R})$ 时，该定理条件(16)与条件(17)也是等价的. 通常称这样的 $A\in\boldsymbol{M}_n(\mathbb{R})$ 为**对角稳定的**（见习题 5）.

还要指出，当 $A\in\mathbb{Z}^{n\times n}$ 时，定理 3 中某些性质显然比起另一些性质更弱些，例如，性质(8)较弱于(5)，性质(12)弱于(6)，性质(13)弱于(14)，等等. 这些事实带来明显的好处. 当我们需要判定矩阵 $A\in\mathbb{Z}^{n\times n}$ 为非奇异 M-矩阵时，可以验证那些较弱的性质. 另一方面，若已知 $A\in\mathbb{Z}^{n\times n}$ 为非奇异 M-矩阵，则可以应用尽可能强的性质去推导进一步的结论.

因为定理 3 条件(5)当 $A\in\boldsymbol{M}_n(\mathbb{R})$ 时重合于 8.1.1 小节定理 3 条件(6)，所以我们有

推论 4　设 $A\in\boldsymbol{M}_n(\mathbb{R})$. 则 $A\in\boldsymbol{P}\cap\mathbb{Z}^{n\times n}$ 当且仅当 A 为非奇异 M-矩阵.

这样一来，对于 $A\in\mathbb{Z}^{n\times n}$，8.1.1 小节定理 3 中条件(2)，(3)，(4)，(5)，(7)，(8)，(9)也是 A 为非奇异 M-矩阵的等价条件（或特征）.

定理 3 的另一个直接推论如下.

推论 5　设 $A\in\mathbb{Z}^{n\times n}$ 为对称矩阵. 则 A 为非奇异 M-矩阵当且仅当 A 为对称

正定的.

通常,将对称正定的矩阵 $A\in\mathbb{Z}^{n\times n}$ 称为 **Stieltjes 矩阵**. 前一推论表明,Stieltjes 矩阵必定为非奇异 M-矩阵. 在历史上,Stieltjes(1887 年)证明了:若 $A\in\boldsymbol{M}_n(\mathbb{R})$ 为对称正定矩阵,且它的非对角元素均为负数,则 $A^{-1}>O$. 从后面讨论看出,Stieltjes 的这个结果为定理 7 的简单推论.

接着,我们讨论任意矩阵 $A\in\boldsymbol{M}_n(\mathbb{R})$ 为非奇异 M-矩阵的充分与必要条件. 在下面定理中,不预先假定 $A\in\mathbb{Z}^{n\times n}$.

定理 6　设 $A\in\boldsymbol{M}_n(\mathbb{R}),n\geqslant 2$. 则下列诸条件分别等价于命题"$A$ 为非奇异 M-矩阵".

(1) 对每一个非负对角矩阵 $D,A+D$ 为逆正的.

(2) $A+\alpha I$ 为逆正的,$\forall\alpha\geqslant 0$.

(3) A 的每一个主子阵为逆正的.

(4) A 的每个一阶,二阶与 n 阶主子阵为逆正的.

证明　设 A 为非奇异 M-矩阵,D 为非负对角矩阵. 则由定理 3(2),有 $\boldsymbol{x}>\boldsymbol{0}$ 使得 $A\boldsymbol{x}>\boldsymbol{0}$,于是,$(A+D)\boldsymbol{x}=A\boldsymbol{x}+D\boldsymbol{x}>\boldsymbol{0}$. 因此,$A+D$ 为非奇异 M-矩阵,由定理 3(11)得$(A+D)$为逆正的. 这样一来,(1)与(2)成立. 还从定理 3 知道,A 的每一个主子阵此时都是非奇异 M-矩阵,因此,(3)与(4)也成立.

现设 $A\in\boldsymbol{M}_n(\mathbb{R})$,且 $A+\alpha I$ 为逆正的,$\forall\alpha\geqslant 0$. 取 $\alpha=0$ 便得 $A^{-1}\geqslant O$. 余下只要验证 $A\in\mathbb{Z}^{n\times n}$ 即可. 事实上,若 A 有元素 $a_{ij}>0(i\neq j)$,则对充分小 $\alpha>0$ 有

$$(I+\alpha A)^{-1}=I-\alpha A+(\alpha A)^2-(\alpha A)^3+\cdots,$$

其中右方第二项 $-\alpha A$ 的(i,j)元素为 $-\alpha a_{ij}<0$ 比起其余各项(i,j)元素的代数和占优势. 因此,$(I+\alpha A)^{-1}$的(i,j)元素为负数,但此与

$$(A+I/\alpha)^{-1}=\alpha(I+\alpha A)^{-1}\geqslant O$$

的事实矛盾. 因为(1)蕴涵(2),所以如果对 $A\in\boldsymbol{M}_n(\mathbb{R})$ 条件(1)成立,则 A 为非奇数 M-矩阵.

显然,条件(3)蕴涵(4). 因此,余下只要证:若 $A\in\boldsymbol{M}_n(\mathbb{R})$ 满足条件(4),则 A 为非奇异 M-矩阵. 此时,首先容易看出,A 的所有对角元素为正数,且 $A^{-1}\geqslant O$. 其次验证 $A\in\mathbb{Z}^{n\times n}$. 令

$$B=\begin{bmatrix}a&b\\c&d\end{bmatrix}$$

为 A 的任意给定的二阶主子阵. 则有 $a>0$ 与 $d>0$,且

$$B^{-1}=(ad-bc)^{-1}\begin{bmatrix}d&-b\\-c&a\end{bmatrix}\geqslant O,$$

因而 $b\leqslant 0$ 与 $c\leqslant 0$. 于是,A 的所有非对角元素为非正的(因为这样元素一定在 A 的某二阶主子阵之中). □

与过去一样,应用矩阵的不可约性质常常可以得到一些加强的结果.

定理 7　设 $A\in\mathbb{Z}^{n\times n}$ 不可约. 则下列诸命题彼此等价:

(1) 存在 $\boldsymbol{x}>\boldsymbol{0}$ 使得 $A\boldsymbol{x}>\boldsymbol{0}$.

(2) 存在 $\boldsymbol{x}>\boldsymbol{0}$ 使得 $A\boldsymbol{x}\geqslant\boldsymbol{0},A\boldsymbol{x}\neq\boldsymbol{0}$.

(3) A 为非奇异 M-矩阵.

(4) $A^{-1}>O$.

证明　(1)⇒(2):显然成立.

(2)⇒(3):设 $\boldsymbol{x}>\boldsymbol{0}$ 满足 $A\boldsymbol{x}\geqslant\boldsymbol{0}$ 与 $A\boldsymbol{x}\neq\boldsymbol{0}$. 将 A 表示为 $A=sI-B,s>0,B\geqslant O$,且令 $T=B/s$. 则 T 非负不可约,且有 $T\boldsymbol{x}\leqslant\boldsymbol{x},T\boldsymbol{x}\neq\boldsymbol{x}$. 若 $T\boldsymbol{x}<\boldsymbol{x}$ 则 $A\boldsymbol{x}>\boldsymbol{0}$ 对 $\boldsymbol{x}>\boldsymbol{0}$ 成立,因而 A 为非奇异 M-矩阵(见定理 3(2)). 若有 $i,j\in\mathcal{N}$ 使得 $(T\boldsymbol{x})_i=x_i$, $(T\boldsymbol{x})_j<x_j$,则由第七章 7.1.1 小节定理 17 推得

$$\rho(T)<\max_k(T\boldsymbol{x})_k/x_k=1,\tag{5}$$

于是,$\rho(B)<s$,即 A 为非奇异 M-矩阵.

(3)⇒(4):设 $A=sI-B,B\geqslant O,s>\rho(B)$. 这时,

$$A^{-1}=(sI-B)^{-1}.$$

由于 $B\in\mathcal{N}_n$ 不可约,从第七章习题 7.1.1 题 11 知道,上式右边矩阵为正矩阵,于是,$A^{-1}>O$.

(4)⇒(1):由定理 3 中(11)与(2)的等价性即得. □

类似于第一章 1.2.2 小节定理 4 后面说明,对于非奇异 M-矩阵,我们有如下结果.

定理 8　设 $A\in\mathbb{Z}^{n\times n}$ 为非奇异 M-矩阵. 则 A 的任意主子阵也是非奇异 M-矩阵. 并且,若 A_{11} 为 A 的某个主子阵,则在 A 内 A_{11} 的 Schur 余量 $[A/A_{11}]$ 也是非奇异 M-矩阵.

证明　本定理的第一个结论为定理 3 性质(6)的直接推论. 为证明第二个结果,我们将 A 写成分块矩阵形式(如必要的话,可进行行与列的同步变换):

$$A=\begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix},$$

于是,$[A/A_{11}]=A_{22}-A_{21}A_{11}^{-1}A_{12}$. 但因 A_{22} 的非对角元素非正,$A_{12}\leqslant O$ 与 $A_{21}\leqslant O$,所以由 $A_{11}^{-1}\geqslant O$ 推出 $[A/A_{11}]$ 的非对角元素皆非正. 根据第一章 1.2.2 小节定理 4 后面说明,$[A/A_{11}]^{-1}$ 为 $A^{-1}\geqslant O$ 的某主子阵,因而 $[A/A_{11}]^{-1}\geqslant O$. 按定理3(11),$[A/A_{11}]$ 为非奇异 M-矩阵. □

非奇异 M-矩阵经常出现在应用差分方法求解偏微分方程所引出的线性代数方程组问题中. 非奇异 M-矩阵的诸多特性可为求解这样的方程组问题提供有力的理论与方法上的依据. 这里仅考虑一个典型的例子.

例 9　应用有限差分方法可将下列求解平面区域 Ω 内的 Dirichlet 问题

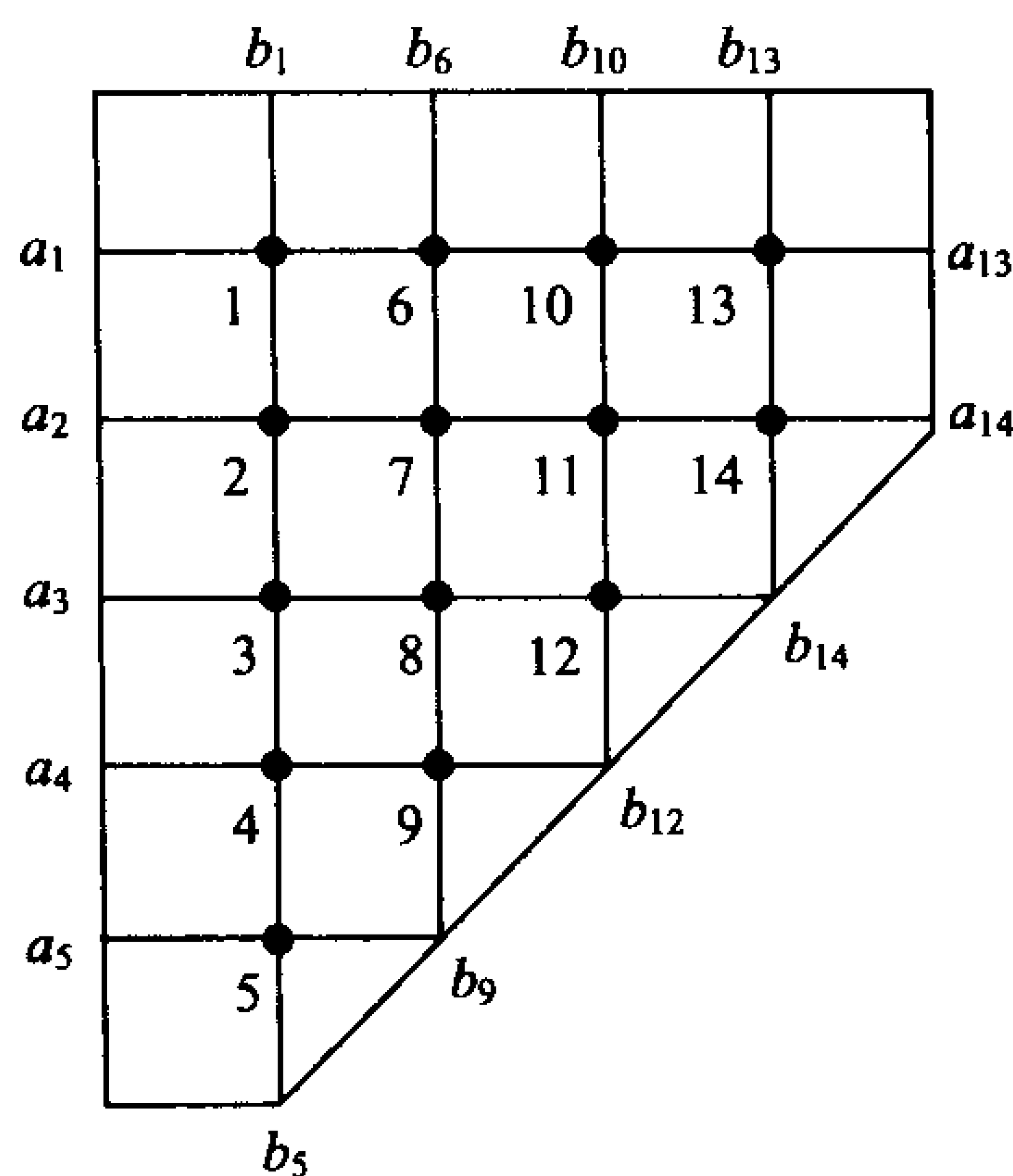


图 8.1

$$\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}=0,\quad (x,y)\in\Omega,$$

$$u(x,y)=g(x,y),\quad (x,y)\in\Gamma \qquad (6)$$

离散为线性代数方程组,式中,Γ 为区域 Ω 的边界.具体地说,若 Ω 与方形网格内部节点与 $u(x,y)$ 在边界 Γ 上的某些值如图 8.1 所示,则相应的线性代数方程组呈如下形式:

$$u_1-\frac{1}{4}u_2\quad -\frac{1}{4}u_6\quad =\frac{1}{4}(a_1+b_1),$$

$$-\frac{1}{4}u_1+u_2-\frac{1}{4}u_3\quad -\frac{1}{4}u_7\quad =\frac{1}{4}a_2,$$

$$\cdots\cdots\cdots$$

$$-\frac{1}{4}u_{11}-\frac{1}{4}u_{13}+u_{14}=\frac{1}{4}(a_{14}+b_{14})$$

其中,u_j 表示 $u(x,y)$ 在内节点 j 处的近似值,a_j 与 b_j 为 $u(x,y)$ 在边界 Γ 上的值.这个方程组具有形式 $A\boldsymbol{u}_0=\boldsymbol{f}_0$,$\boldsymbol{u}_0=(u_1,u_2\cdots,u_{14})^{\mathrm{T}}$,$\boldsymbol{f}_0=\left(\frac{1}{4}(a_1+b_1),\frac{1}{4}a_2,\cdots,\frac{1}{4}(a_{14}+b_{14})\right)^{\mathrm{T}}\in\mathbb{R}^{14}$,而 $A\in\mathbb{Z}^{14\times 14}$ 为不可约的非奇异 M-矩阵,因为对 $\boldsymbol{e}=(1,1,\cdots,1)^{\mathrm{T}}\in\mathbb{R}^{14}$,$A\boldsymbol{e}\geqslant\boldsymbol{0}$ 与 $A\boldsymbol{e}\neq\boldsymbol{0}$.于是,按定理 7,$A^{-1}$ 存在且 $A^{-1}>O$.这表明上述方程组有唯一解 $\boldsymbol{u}_0$,且当 $\boldsymbol{f}_1\geqslant\boldsymbol{f}_0$ 时,$\boldsymbol{u}_1=A^{-1}\boldsymbol{f}_1\geqslant A^{-1}\boldsymbol{f}_0=\boldsymbol{u}_0$(若还有 $\boldsymbol{f}_0\neq\boldsymbol{f}_1$,则 $\boldsymbol{u}_1>\boldsymbol{u}_0$). □

习题 8.1.2

1. 试证:设 $A\in\boldsymbol{M}_n(\mathbb{R})$.则下列三个条件等价:

(1) A 为半正的,即有 $\boldsymbol{x}>\boldsymbol{0}$ 使得 $A\boldsymbol{x}>\boldsymbol{0}$.

(2) 有 $\boldsymbol{x}\geqslant\boldsymbol{0}$ 使得 $A\boldsymbol{x}>\boldsymbol{0}$.

(3) 有正对角矩阵 D 使得 AD 所有的行和为正数.

2. 设 $A\in\boldsymbol{M}_n(\mathbb{R})$.试证:若存在置换矩阵 P 使得 $B\equiv PAP^{\mathrm{T}}$ 满足条件:有 $\boldsymbol{x}>\boldsymbol{0}$ 使得 $B\boldsymbol{x}\geqslant\boldsymbol{0}$ 且

$$\sum_{j=1}^{i}b_{ij}x_j>0,\qquad i=1,2\cdots,n,$$

则 A 为半正的.

3. 设 $A\in\boldsymbol{M}_n(\mathbb{R})$.试证下列两个条件彼此等价:

(1) A 为**广义正型的**,即存在 $\boldsymbol{x}>\boldsymbol{0}$ 使得 $\boldsymbol{y}=A\boldsymbol{x}\geqslant\boldsymbol{0}$ 与 $\boldsymbol{y}\neq\boldsymbol{0}$,且若 $y_{i_0}=0$,则有 $\mathcal{N}$ 的序列 $i_1,\cdots,i_r$ 使得

$$a_{i_{j-1}i_j}\neq 0,\qquad j=1,\cdots,r\quad 与\quad y_{i_r}\neq 0.$$

(2) 存在 $\boldsymbol{x}>\boldsymbol{0}$ 使得 $\boldsymbol{y}=A\boldsymbol{x}\geqslant\boldsymbol{0}$ 与 $\boldsymbol{y}\neq\boldsymbol{0}$,并且由下列确定的矩阵 $\hat{A}=(\hat{a}_{ij})\in\mathcal{N}_n$ 不可约:

$$\hat{a}_{ij}=\begin{cases}1, & \text{如果 } a_{ij}\neq 0 \text{ 或 } y_i\neq 0,\\ 0, & \text{其他情形}.\end{cases}$$

4. 设 $A\in \boldsymbol{M}_n(\mathbb{R})$. 证明下列诸命题彼此等价：

(1) A 逆正，即 A 可逆且 $A^{-1}\geqslant O$.

(2) A 单调，即 $A\boldsymbol{x}\geqslant A\boldsymbol{y}$ 蕴涵 $\boldsymbol{x}\geqslant \boldsymbol{y}$.

(3) 存在逆正矩阵 B_1 与 B_2 使得 $B_1\leqslant A\leqslant B_2$.

(4) 存在逆正矩阵 B 使得 $B\geqslant A$ 且 $\rho(I-B^{-1}A)<1$.

(5) 存在逆正矩阵 B 使得 $B\geqslant A$ 且 A 满足题 1 中(1),(2)或(3).

(6) 存在逆正矩阵 B 与非奇异 M-矩阵 C 使得 $B\geqslant A$ 与 $A=BC$.

(7) 存在逆正矩阵 B 与非奇异 M-矩阵 C 使得 $A=BC$.

(8) A 有收敛的规则分裂，即 A 可以表达为 $A=P-Q$，其中 $P^{-1}\geqslant O, Q\geqslant O$ 且 $\rho(P^{-1}Q)<1$.

(9) A 有**收敛的弱规则分裂**，即 A 可以表达为 $A=P-Q$，其中 $P^{-1}\geqslant O, P^{-1}Q\geqslant O$ 且 $\rho(P^{-1}Q)<1$.

5. 证明：当 $A\in \boldsymbol{M}_n(\mathbb{R})$ 时，定理 3 中条件(16)与(17)也是等价的. 并且，它们还等价于下列命题：

"对每一个非零的对称正半定矩阵 Q，矩阵 QA 有一个正的对角元素".

6. 设 $A\in \boldsymbol{M}_n(\mathbb{C})$. 定义 $\mathcal{M}(A)=(m_{ij})\in \boldsymbol{Z}^{n\times n}$ 如下：

$$m_{ii}=|a_{ii}|, m_{ij}=-|a_{ij}|, \qquad \forall i\neq j, 1\leqslant i,j\leqslant n,$$

我们称之为 A 的**比较矩阵**(comparison matrix). 试证：若 $A\in \boldsymbol{M}_n(\mathbb{R})$ 的所有对角元素 $a_{ii}>0$，则 $\mathcal{M}(A)$ 为非奇异 M-矩阵当且仅当存在 $\boldsymbol{x}>\boldsymbol{0}$ 使得对所有 n 阶符号矩阵 S，$SAS\boldsymbol{x}>\boldsymbol{0}$.

7. 设 $A=(a_{ij})\in \boldsymbol{M}_n(\mathbb{R})$. 证明下列诸命题彼此等价：

(1) 存在 $\boldsymbol{x}>\boldsymbol{0}$ 使得对所有 n 阶符号矩阵 S，$SAS\boldsymbol{x}>\boldsymbol{0}$.

(2) 所有 $a_{ii}>0$，且存在正对角矩阵 $D=\operatorname{diag}[d_1,\cdots,d_n]$ 使得 AD 严格对角占优，即 $a_{ii}d_i>\sum_{j\neq i}|a_{ij}|d_j, \forall i\in \mathcal{N}$.

(3) 所有 $a_{ii}>0$，且存在正对角矩阵 $E=\operatorname{diag}[e_1,\cdots,e_n]$ 使得 $E^{-1}AE$ 严格对角占优，即

$$a_{ii}>\frac{1}{e_i}\sum_{j\neq i}|a_{ij}|e_j, \qquad \forall i\in \mathcal{N}.$$

(4) 所有 $a_{ii}>0$，且存在正对角矩阵 $D=\operatorname{diag}[d_1,\cdots,d_n]$ 使得 AD 下半严格对角占优(见第六章习题 6.1.2 题 6)，即

$$a_{ii}d_i\geqslant \sum_{j\neq i}|a_{ij}|d_j, \qquad \forall i\in \mathcal{N},$$

$$a_{ii}d_i>\sum_{j=1}^{i-1}|a_{ij}|d_j, \qquad i=2,\cdots,n.$$

8.1.3 G-函数与非奇异 M-矩阵

对于 $A\in \boldsymbol{M}_n(\mathbb{R})$，我们在 8.1.1 小节与 8.1.2 小节中(包括习题)介绍了各种条件与性质. 这些都以习题 8.1.2 题 4 或题 7 中介绍的两组等价性质为基础，因为可以证明其他性质都可以从 A 的逆正性(即单调性)或从 A 广义严格对角占优且

所有对角元素为正数的性质推导出来(见习题 1). 并且对 $A\in\mathbb{Z}^{n\times n}$,这些条件均等价于"$A$ 为非奇异 M-矩阵"(见习题 2). 本小节专门讨论严格对角占优矩阵及其推广(见第六章 6.2 节)与非奇异 M-矩阵之间的关系.

首先引入 G-函数概念.

在第六章 6.1 节与 6.2 节中,我们已经建立一些准则用以判断一个复矩阵 $A\in\boldsymbol{M}_n(\mathbb{C})$的非奇异性,其中最基本的是 Lévy-Desplanques 准则(第六章 6.1.1 小节定理 2)及其推论(第六章 6.1.1 小节推论 3). 对于不可约矩阵 $A\in\boldsymbol{M}_n(\mathbb{C})$,则有 Taussky 定理(第六章 6.1.2 小节定理 8). 同时,我们还证明了另外一些判定矩阵非奇异的准则,例如,第六章 6.2.1 小节定理 3 等. 在这些结果中形如

$$|a_{ii}|>R_i(A),\qquad i=1,\cdots,n \tag{1}$$

条件里的 $R_i(A)$分别被代以 $R_i^x(A)$, $C_i(A)$, $R_{i,\alpha p}(A)C_{i,(1-\alpha)q}^{1-\alpha}(A)$ $(0<\alpha<1, p\geqslant 1, q\geqslant 1, p^{-1}+q^{-1}=1)$,以及 $\alpha R_{i,\alpha p}(A)+(1-\alpha)C_{i,(1-\alpha)q}(A)$等. 这些 n-组数,譬如$(R_1(A), R_2(A),\cdots,R_n(A))$,只与 A 的非对角元素的模有关,而与 A 的对角元素无关. 因此,若 $A,B\in\boldsymbol{M}_n(\mathbb{C})$,只要$|a_{ij}|=|b_{ij}|$, $\forall i\neq j$,便有

$$\begin{aligned}
&(R_1(A),\cdots,R_n(A))=(R_1(B),\cdots,R_n(B)),\\
&(R_1^x(A),\cdots,R_n^x(A))=(R_1^x(B),\cdots,R_n^x(B)),\qquad \forall\boldsymbol{x}>\boldsymbol{0},\\
&(R_{1,\alpha p}^{\alpha}(A)C_{1,(1-\alpha)q}^{1-\alpha}(A),\cdots,R_{n,\alpha p}^{\alpha}(A)C_{n,(1-\alpha)q}^{1-\alpha}(A))\\
&=(R_{1,\alpha p}^{\alpha}(B)C_{1,(1-\alpha)q}^{1-\alpha}(B),\cdots,R_{n,\alpha p}^{\alpha}(B)C_{n,(1-\alpha)q}^{1-\alpha}(B)).
\end{aligned}$$

这十分自然地导出 G-函数的概念.

设 $f_i:\boldsymbol{M}_n(\mathbb{C})\to\mathbb{R}_+$ 只与 $\boldsymbol{M}_n(\mathbb{C})$中矩阵的非对角元素的模有关. 用 $\mathscr{C}_n(n\geqslant 2)$表示这些 f_i 组成的 n-组数$(f_1,\cdots,f_n)$的集合.

定义 1 我们称 $f=(f_1,\cdots,f_n)\in\mathscr{C}_n$ 为一个 **G-函数**,假如满足条件

$$|a_{ii}|>f_i(A),\qquad i=1,\cdots,n \tag{2}$$

的每一个矩阵 $A\in\boldsymbol{M}_n(\mathbb{C})$为非奇异的. 等价地,$f=(f_1,\cdots,f_n)\in\mathscr{C}_n$ 为一个 G-函数,如果任意矩阵 $A\in\boldsymbol{M}_n(\mathbb{C})$的所有特征值位于如下 n 个圆盘的并集内:

$$\{z\in\mathbb{C}: |z-a_{ii}|\leqslant f_i(A)\},\qquad i=1,\cdots,n. \tag{3}$$

今后用 $\mathscr{G}_n$ 表示 $\mathscr{C}_n$ 中所有 G-函数的集合. 这样,前面介绍的 $R=(R_1,\cdots,R_n)$等均为 G-函数. 有趣的是,虽然可以找出各式各样的 G-函数,但下面的 Ky Fan(樊畿)定理指出,对不可约矩阵情形,最原始的 Lévy-Desplanques 准则仍然是"最佳"的判定矩阵非奇异性的准则,当然,其中允许条件(1)中的 $R_i(A)$被 $R_i^x(A)$所代替,$i=1,\cdots,n$,这里 $\boldsymbol{x}=(x_1,\cdots,x_n)^{\mathrm{T}}$ 为任意的正向量. 为此先证明一个简单的引理,它是第六章 6.1.2 小节推论 5 的一种推广.

引理 2 设 $A=(a_{ij})\in\boldsymbol{M}_n(\mathbb{C})$ $(n\geqslant 2)$不可约,且 $f=(f_1,\cdots,f_n)\in\mathscr{G}_n$. 则有 $f_i(A)>0$, $i=1,\cdots,n$.

证明 令 $B=(b_{ij})\in\boldsymbol{M}_n(\mathbb{C})$)满足 $b_{ii}=f_i(A)$, $b_{ij}=-|a_{ij}|$, $\forall i\neq j$, $1\leqslant i,j\leqslant$

n. 按定义 1, $f_i(A)=f_i(B)$, $i=1,\cdots,n$. 显然, $B\in\mathbb{Z}^{n\times n}$. 将 B 表达为 $B=sI-Q$, $s>\max\limits_i f_i(A)$, $Q\geqslant O$, 则有 $s\geqslant\rho(Q)$, 这是因为对任意给定的 $\varepsilon>0$, $B+\varepsilon_1 I=(s+\varepsilon_1)I-Q$ 满足 $f_i(A)+\varepsilon_1>f_i(B+\varepsilon_1 I)=f_i(B)$, $\forall\,\varepsilon_1\geqslant\varepsilon$ 与 $i=1,\cdots,n$, 所以由定义 1, $B+\varepsilon I$ 为非奇异的, 因而按 8.1.2 小节定理 3(18) 与习题 8.1.1 题 2, $B+\varepsilon I$ 为非奇异 M-矩阵. 于是, $s+\varepsilon>\rho(Q)$, $\forall\,\varepsilon>0$, 从而 $s\geqslant\rho(Q)$, 即 B 为 M-矩阵. 另一方面, 由于 $Q\in\mathcal{N}_n$ 不可约, 根据 Perron-Frobenius 定理, 半径 $q(Q)$ 为 Q 的单重特征值, 且 Q 的任意真正主子阵 $\widetilde{Q}$ 的谱半径 $\rho(\widetilde{Q})<\rho(Q)$, 故 $\widetilde{B}\equiv s\widetilde{I}-\widetilde{Q}$ 满足 $s\geqslant\rho(Q)>\rho(\widetilde{Q})$, 因而 $\widetilde{B}$ 为非奇异 M-矩阵. 这表明 $\widetilde{B}$ 因而 B 的对角元素必定为正数, 即有 $f_i(A)>0$, $i=1,\cdots,n$. □

定理 3(Ky Fan[5]) 设 $A=(a_{ij})\in\boldsymbol{M}_n(\mathbb{C})$ $(n\geqslant 2)$ 不可约, 且 $f=(f_1,\cdots,f_n)\in\mathcal{G}_n$. 则存在(与 A 有关的) $\boldsymbol{x}=(x_1,\cdots,x_n)^{\mathrm{T}}>\boldsymbol{0}$ 使得

$$f_i(A)\geqslant R_i^x(A),\qquad i=1,\cdots,n. \tag{4}$$

证明 根据引理 2, $f_i(A)>0$, $i=1,\cdots,n$. 定义 $\widetilde{A}=(\tilde{a}_{ij})\in\boldsymbol{M}_n(\mathbb{R})$: $\tilde{a}_{ii}=0$ $(1\leqslant i\leqslant n)$ 与 $\tilde{a}_{ij}=|a_{ij}|/f_i(A)$, $\forall\,i\neq j$. 显然, $\widetilde{A}\in\mathcal{N}_n$ 不可约. 按 Perron-Frobenius 定理, $\lambda=\rho(\widetilde{A})>0$ 为 $\widetilde{A}$ 的特征值, 且有正的 Perron 向量 $\boldsymbol{x}=(x_1,\cdots,x_n)^{\mathrm{T}}$ 使得

$$\lambda x_i=\sum_{j\neq i}|a_{ij}|x_j/f_i(A),\qquad i=1,\cdots,n, \tag{5}$$

因而

$$0<\lambda=\frac{1}{x_i}\sum_{j\neq i}|a_{ij}|x_j/f_i(A)=R_i^x(A)/f_i(A),\qquad i=1,\cdots,n. \tag{6}$$

现讨论奇异矩阵 $B=\widetilde{A}-\lambda I=(b_{ij})$. 显然, $b_{ij}=\tilde{a}_{ij}$, $\forall\,i\neq j$. 由于 B 奇异, 故至少有一个 $i\in\mathcal{N}$, 使得 $|b_{ii}|\leqslant 1$. 事实上, 若 $|b_{ii}|>1$ $(1\leqslant i\leqslant n)$, 则 $\widetilde{B}\equiv\mathrm{diag}[f_1(A),\cdots,f_n(A)]B\equiv(\tilde{b}_{ij})$ 满足 $|\tilde{b}_{ii}|=f_i(A)|b_{ii}|>f_i(A)=f_i(\widetilde{B})$, $\forall\,i\in\mathcal{N}$, 因为 $\widetilde{B}$ 的非对角元素 $\tilde{b}_{ij}=f_i(A)b_{ij}=f_i(A)\tilde{a}_{ij}=|a_{ij}|$, $\forall\,i\neq j$. 因此, $\widetilde{B}$(因而 B) 非奇异, 推出矛盾. 但是 $|b_{ii}|=\lambda$, $\forall\,i\in\mathcal{N}$, 于是, $\lambda\leqslant 1$. 应用(6)式便得(4)式. □

定理 3 有如下三个直接的推论.

推论 4 设 $A=(a_{ij})\in\boldsymbol{M}_n(\mathbb{C})$ 不可约, 且 $f=(f_1,\cdots,f_n)\in\mathcal{G}_n$. 若

$$|a_{ii}|\geqslant f_i(A),\qquad i=1,2,\cdots,n, \tag{7}$$

其中至少有一个严格不等式成立, 则 $\det A\neq 0$. 特别地, 如还有 $a_{ii}>0$ $(1\leqslant i\leqslant n)$, 则 $\det A>0$.

推论 5 设 $A=(a_{ij})\in\boldsymbol{M}_n(\mathbb{C})$ $(n\geqslant 3)$ 不可约, 且 $2\leqslant k\leqslant n-1$. 若对某 $f=(f_1,\cdots,f_n)\in\mathcal{G}_n$ 有

$$\prod_{i\in\mathcal{S}}|a_{ii}|\geqslant\prod_{i\in\mathcal{S}}f_i(A),\qquad\forall\,\mathcal{S}, \tag{8}$$

式中, $\mathcal{S}\subset\mathcal{N}$ 涵义如同第六章 6.2.2 小节定理 1, 则 $\mathrm{rank}A\geqslant n-k+1$.

这是第六章 6.2.2 小节定理 3 的推广.

推论 6 设 $A=(a_{ij})\in \boldsymbol{M}_n(\mathbb{C})(n\geqslant 2)$不可约. 若对某 $f=(f_1,\cdots,f_n)\in\mathscr{G}_n$,

$$\sum_{i\in v}|a_{ii}|\geqslant\sum_{i\in v}f_i(A) \tag{9}$$

对 $\Gamma(A)$中所有简单回路 v 都成立,且至少对 $\Gamma(A)$的一个简单回路(9)式中严格不等式成立,则 $\det A\neq 0$.

这是第六章 6.2.2 小节定理 6 的推广.

定理 3 可以推广到可约矩阵的情形. 考虑 $A\in\boldsymbol{M}_n(\mathbb{C})(n\geqslant 2)$的法式:

$$PAP^{\mathrm{T}}=\begin{bmatrix}A_{11} & A_{12} & \cdots & A_{1m}\\ & A_{22} & \cdots & A_{2m}\\ & 0 & \ddots & \vdots\\ & & & A_{mm}\end{bmatrix}, \tag{10}$$

其中块对角线上每块 $A_{jj}(1\leqslant j\leqslant m)$或为不可约矩阵,或为一阶零矩阵(见第七章 7.2.1 小节). 令 A_{jj}的下标集合为 $\mathscr{S}_j(1\leqslant j\leqslant m)$,则各 $\mathscr{S}_j$ 之间彼此不交且 $\bigcup_{j=1}^{m}\mathscr{S}_j=\mathscr{N}=\{1,2,\cdots,n\}$. 用$\langle i\rangle$表示包含 i 的某个下标集 $\mathscr{S}_k$ 且对每一个正向量 $\boldsymbol{x}=(x_1,\cdots,x_n)^{\mathrm{T}}$,定义

$$\hat{R}_i^{x}(A)=\frac{1}{x_i}\sum_{\substack{j\in\langle i\rangle\\ j\neq i}}|a_{ij}|x_j,\qquad i=1,\cdots,n. \tag{11}$$

(若$\langle i\rangle$仅含有 $\mathscr{N}$的唯一元素,则规定 $\hat{R}_i^{x}(A)=0$)

容易验证,对任意正向量 $\boldsymbol{x}=(x_1,\cdots,x_n)^{\mathrm{T}}$,$\hat{R}^{x}=(\hat{R}_1^{x},\hat{R}_2^{x},\cdots,\hat{R}_n^{x})\in\mathscr{C}_n$ 皆为G-函数,并且,

$$\hat{R}_i^{x}(A)\leqslant R_i^{x}(A),\qquad \forall A\in\boldsymbol{M}_n(\mathbb{C})\text{ 与 }\forall i\in\mathscr{N}$$

(见习题 3).

定理 7(Carlson-Varga[8]) 设 $A=(a_{ij})\in\boldsymbol{M}_n(\mathbb{C})(n\geqslant 2)$,且 $f=(f_1,\cdots,f_n)\in\mathscr{G}_n$. 则存在(与 A 有关的)$\boldsymbol{x}=(x_1,\cdots,x_n)^{\mathrm{T}}>\boldsymbol{0}$ 使得

$$f_i(A)\geqslant\hat{R}_i^{x}(A),\qquad i=1,\cdots,n. \tag{12}$$

证明 若 A 不可约,则显然有 $\hat{R}_i^{x}(A)=R_i^{x}(A)$,$\forall\,\boldsymbol{x}>\boldsymbol{0}$ 与$\forall\, i\in\mathscr{N}$,因而由定理 3 知道(12)式成立. 余下只要讨论 A 为可约的情形. 假定 A 有可约法式(10),我们定义 $\mathscr{M}^f(A)=(m_{ij})\in Z^{n\times n}$:$m_{ii}=f_i(A)(1\leqslant i\leqslant n)$与 $m_{ij}=-|a_{ij}|$,$\forall\, i\neq j$. 根据引理 2 的证明,$\mathscr{M}^f(A)$为 n 阶M-矩阵,且它的所有特征值的实部非负. 对应于(10)式中每一个 A_{kk},有一个与 A_{kk}同阶的矩阵 $\mathscr{M}_{kk}^f(A)$:

$$(\mathscr{M}_{kk}^f(A))_{ij}=\begin{cases}f_i(A), & \text{如果 } i=j\in\mathscr{S}_k,\\ -|a_{ij}|, & \text{如果 } i\neq j;i,j\in\mathscr{S}_k.\end{cases} \tag{13}$$

显然,$\mathscr{M}_{kk}^f(A)(1\leqslant k\leqslant m)$或为不可约的或为一阶矩阵(其唯一元素为 $f_i(A)$,假如 $\mathscr{S}_k=\{i\}$). 将 $\mathscr{M}_{kk}^f(A)$表示为

$$\mathscr{M}_{kk}^{f}(A)=\alpha_k I-B_k,\qquad k=1,\cdots,m$$

式中，$\alpha_k=\max\limits_{i\in\mathscr{S}_k}f_i(A)$，则 B_k 或为非负不可约矩阵或为一阶零矩阵. 当 B_k 为非负不可约矩阵时，B_k 有正特征值 $\rho(B_k)$ 与对应的正特征向量 $\boldsymbol{x}^{(k)}$. 此时，$\mathscr{M}_{kk}^{f}(A)$ 有实特征值 $\alpha_k-\rho(B_k)$，它显然也是 $\mathscr{M}^f(A)$ 的实特征值，因而 $\alpha_k-\rho(B_k)\geqslant 0$. 当 B_k 为一阶零矩阵时，$\rho(B_k)=0$ 为 B_k 的特征值，它对应有正特征向量 $\boldsymbol{x}^{(k)}$. 现令

$$g_i(A)=f_i(A)-\lambda_k(A),\qquad i\in\mathscr{S}_k,k=1,\cdots,m,$$

其中，$\lambda_k(A)=\alpha_k-\rho(B_k)\geqslant 0$. 特别地当 $\mathscr{S}_k=\{i\}$ 时，$g_i(A)=0$. 考虑矩阵 $\mathscr{M}_{kk}^{f}(A)-\lambda_k(A)I$. 它是奇异的，并且，

$$(\mathscr{M}_{kk}^{f}(A)-\lambda_k(A)I)\boldsymbol{x}^{(k)}=(\alpha_k-\rho(B_k)-\lambda_k(A))\boldsymbol{x}^{(k)}=\boldsymbol{0}.$$

于是，

$$g_i(A)x_i^{(k)}=\sum_{\substack{j\in\langle i\rangle\\ j\neq i}}|a_{ij}|x_j^{(k)},\qquad \forall i\in\mathscr{S}_k,k=1,\cdots,m.$$

按前面规定，当 $\mathscr{S}_k=\{i\}$ 时，上式右端为零. 由于 $0\leqslant g_i(A)\leqslant f_i(A)(1\leqslant i\leqslant n)$，我们有

$$f_i(A)\geqslant\frac{1}{x_i^{(k)}}\sum_{\substack{j\in\mathscr{S}_k\\ j\neq i}}|a_{ij}|x_j^{(k)},\qquad \forall i\in\mathscr{S}_k,k=1,\cdots,m.$$

$\boldsymbol{x}^{(k)}(1\leqslant k\leqslant m)$ 的所有分量组成一个正向量 $\boldsymbol{x}\in\mathbb{R}^n$，对此 $\boldsymbol{x}$，(12)式成立. □

应用定理 7，我们来证明 Shemesh 定理的一种推广(见第六章 6.2.2 小节定理 1).

定理 8 设 $A=(a_{ij})\in\boldsymbol{M}_n(\mathbb{C})(n\geqslant 2)$，$f=(f_1,\cdots,f_n)\in\mathscr{G}_n$，且 $1\leqslant k\leqslant n-1$. 若对 $\mathscr{N}$ 的任意 $k+1$ 个不同元素组成的子集 $\mathscr{S}$ 有

$$\sum_{i\in\mathscr{S}}|a_{ii}|>\sum_{i\in\mathscr{S}}f_i(A),\tag{14}$$

则 $\operatorname{rank}A=n-k+1$.

证明 当 A 不可约时，结论由第六章 6.2.2 小节定理 1 与这里定理 3 直接推出. 余下考虑 A 为可约的情形. 设 A 有法式(10). 由第六章 6.2.2 小节定理 1 证明看出，我们仅仅需要验证两个事实：首先，当 A 可约时，如果 $1\leqslant p<k$，且

$$|a_{ii}|>f_i(A),\qquad i=p+1,\cdots,n,\tag{15}$$

则有 $\operatorname{rank}A\geqslant n-p$；其次，当 A 可约时，如果 $p=k$，且

$$|a_{ii}||a_{jj}|>f_i(A)f_j(A),\qquad i\neq j;i,j=k,\cdots,n,\tag{16}$$

则有 $\operatorname{rank}A\geqslant n-k+1$. 容易看出，前一事实可由定理 7 直接推出. 为验证第二个事实，我们定义矩阵 $\widetilde{A}$ 为

$$\widetilde{A}=P^{\mathrm{T}}\operatorname{diag}[A_{11},A_{22},\cdots,A_{mm}]P,\tag{17}$$

这里置换矩阵 P 以及诸 A_{jj} 都与(10)式一致. 显然，$\operatorname{rank}A\geqslant\operatorname{rank}\widetilde{A}$，且对任一正向量 $\boldsymbol{x}=(x_1,\cdots,x_n)^{\mathrm{T}}$，$\hat{R}_i^x(A)=R_i^x(\widetilde{A})$，$1\leqslant i\leqslant n$. 应用定理 7，有 $\boldsymbol{x}=(x_1,\cdots,x_n)^{\mathrm{T}}>\boldsymbol{0}$ 使得(12)式成立. 因此，(16)式蕴涵

$$|a_{ii}||a_{jj}| > \hat{R}_i^x(A)\hat{R}_j^x(A) = R_i^x(\widetilde{A})R_j^x(\widetilde{A}), \qquad i \neq j; i,j = k,\cdots,n.$$

再应用第六章 6.2.1 小节定理 1，矩阵 $\widetilde{A}$ 有一个 $n-k+1$ 阶的主子阵为满秩的，于是 $\mathrm{rank}A \geqslant \mathrm{rank}\widetilde{A} \geqslant n-k+1$. □

我们指出，在 Shemesh 的工作中，上述定理中的 $f_i(A) = R_i^{\alpha}(A)C_i^{1-\alpha}(A)$，$i=1,\cdots,n$，$0 \leqslant \alpha \leqslant 1$. 因此，定理 8 比起 Shemesh 结果更为一般.

在定理 8 的结果中，如取 $k=1$，我们有

推论 9　设 $A=(a_{ij}) \in \boldsymbol{M}_n(\mathbb{C})(n \geqslant 2)$，$f=(f_1,\cdots,f_n) \in \mathscr{G}_n$. 若

$$|a_{ii}||a_{jj}| > f_i(A)f_j(A), \qquad i \neq j; i,j = 1,\cdots,n, \tag{18}$$

则 $\det A \neq 0$.

显然，这个推论比起第六章 6.2.1 小节定理 1 与定理 5 来包含着更为一般的结果. 从这个推论出发，也可得到矩阵特征值的排除区域.

现在讨论 G 函数与 M-矩阵之间的关系. 我们在习题 8.1.2 题 6 中已经引入比较矩阵的定义. 显然，对 $A=(a_{ij}) \in \boldsymbol{M}_n(\mathbb{C})$，它的比较矩阵 $\mathscr{M}(A)=(m_{ij}) \in \mathbb{Z}^{n\times n}$，且对任意 $f=(f_1,\cdots,f_n) \in \mathscr{G}_n$，$f_i(A) = f_i(\mathscr{M}(A))$，$i=1,\cdots,n$.

定理 10　设 $A=(a_{ij}) \in \boldsymbol{M}_n(\mathbb{C})(n \geqslant 2)$. 则下列命题互相等价：

(1) $\mathscr{M}(A)$ 为非奇异 M-矩阵.

(2) 存在 $f=(f_1,\cdots,f_n) \in \mathscr{G}_n$ 使得

$$|a_{ii}| > f_i(A), \qquad i = 1,\cdots,n.$$

(3) 存在 $f=(f_1,\cdots,f_n) \in \mathscr{G}_n$ 使得

$$|a_{ii}||a_{jj}| > f_i(A)f_j(A), \qquad i \neq j; i,j = 1,\cdots,n.$$

(4) 所有 $a_{ii} \neq 0$，且存在 $f=(f_1,\cdots,f_n) \in \mathscr{G}_n$ 使得

$$\prod_{i\in v} |a_{ii}| > \prod_{i\in v} f_i(A), \tag{19}$$

其中，v 为 A 的有向图 $\Gamma(A)$ 的任意简单回路.

证明　(1)⇒(2)：由 8.1.2 小节定理 3 知道，存在 $\boldsymbol{x}=(x_1,\cdots,x_n)^{\mathrm{T}} > \boldsymbol{0}$ 使得 $\mathscr{M}(A)\boldsymbol{x} > \boldsymbol{0}$，即有

$$|m_{ii}| > \frac{1}{x_i}\sum_{j\neq i}(-m_{ij})x_j, \qquad i = 1,\cdots,n.$$

由于 $-m_{ij} = |a_{ij}|$，$\forall i \neq j$，上式等同于 $|a_{ii}| > R_i^x(A)$，$i=1,\cdots,n$. 取 $f=(f_1,\cdots,f_n)=(R_1^x,\cdots,R_n^x) \in \mathscr{G}_n$ 便知(2)成立.

(2)⇒(3)与(3)⇒(4)都是显然的.

(4)⇒(1)：容易证明，满足条件(4)的矩阵 $A \in \boldsymbol{M}_n(\mathbb{C})$ 必为非奇异的(见习题 4)，因而 $\det\mathscr{M}(A) \neq 0$. 对任意实数 $\varepsilon \geqslant 0$，$\mathscr{M}(A)+\varepsilon I$ 的所有对角元素为正数，且由(19)式推出

$$\prod_{i\in v}(|a_{ii}|+\varepsilon) > \prod_{i\in v} f_i(A), \qquad \forall v,$$

于是 $\mathscr{M}(A)+\varepsilon I\in\mathbb{Z}^{n\times n}$ 非奇异，$\forall\varepsilon\geqslant 0$. 因此按习题 8.1.1 题 2 与 8.1.2 小节定理 3，$\mathscr{M}(A)$ 为非奇异 M-矩阵. □

习题 8.1.3

1. 设 $A\in\boldsymbol{M}_n(\mathbb{R})$，用下列固定的字母表示关于 A 的某些条件：

(A) A 的所有主子式为正数.

(B) A 的所有 k 阶主子式之和为正数，$k=1,\cdots,n$.

(C)：$\det A\neq 0$ 且 A 的所有主子式为非负数.

(D)：$\det(A+\varepsilon I)\neq 0$，$\forall\varepsilon\geqslant 0$.

(E)：A 的所有前主子式为正数.

(F)：存在置换矩阵 P 使得 PAP^{T} 满足条件(E).

(G)：A 为正稳定的，即 A 的每一个特征值的实部为正的.

(H)：A 为对角稳定的，即存在正对角矩阵 D 使得 $AD+DA^{\mathrm{T}}$ 为对称正定的.

(I)：A 为半正的.

(J)：存在 $\boldsymbol{x}>\boldsymbol{0}$ 使得 $A\boldsymbol{x}\geqslant\boldsymbol{0}$ 且 $\sum_{j=1}^{i}a_{ij}x_j>0$，$i=1,\cdots,n$.

(K)：存在置换矩阵 P 使得 PAP^{T} 满足条件(J).

(L)：A 为广义正型的(见习题 8.1.2 题 3).

(M)：所有 $a_{ii}>0$，且 A 为广义严格对角占优的.

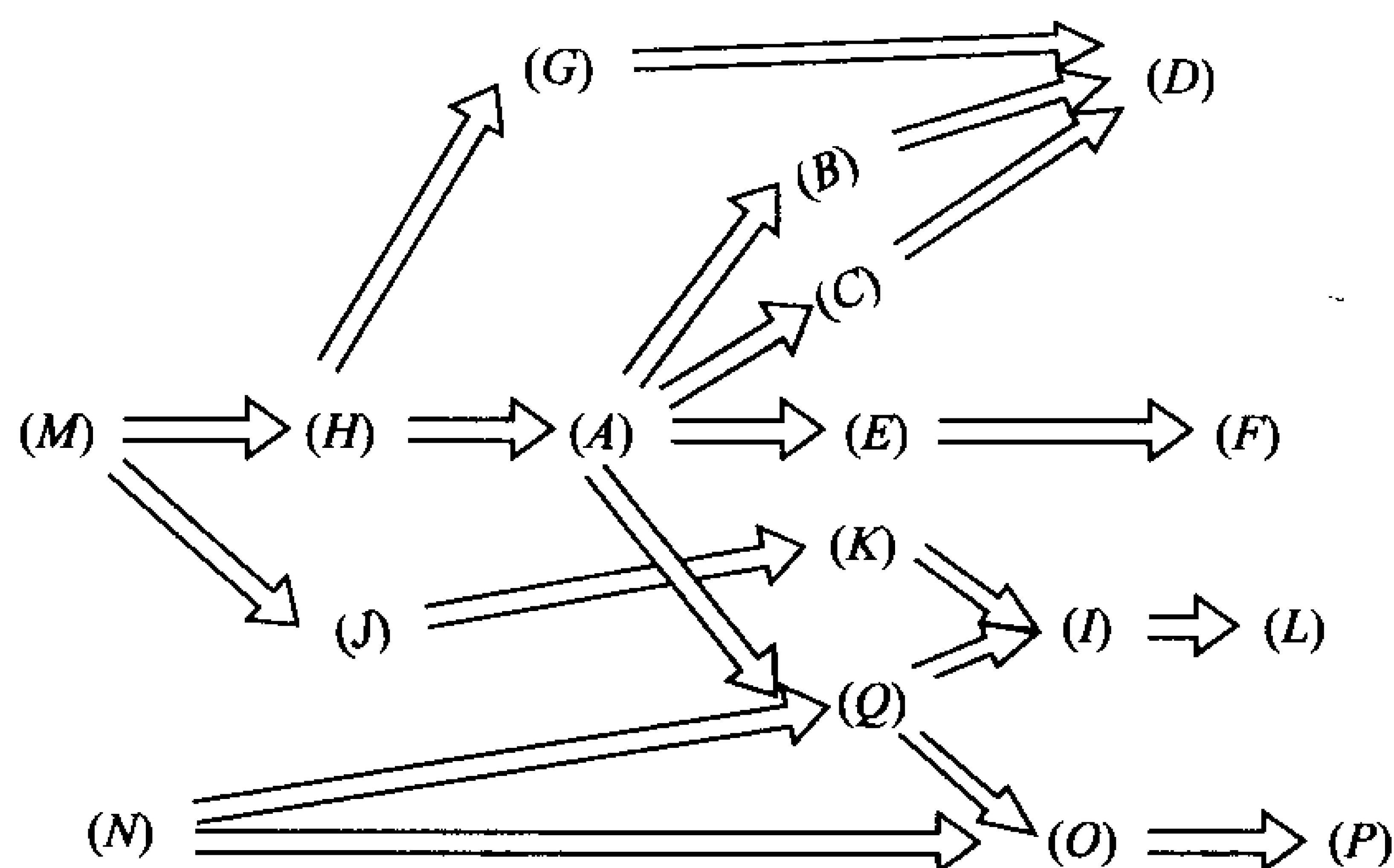
(N)：A 为逆正的.

(O)：A 的每一个弱规则分裂为收敛的，即若 A 有表达式 $A=M-N$，$M^{-1}\geqslant O$，$M^{-1}N\geqslant O$，则 $\rho(M^{-1}N)<1$.

(P)：A 的每一个规则分裂为收敛的，即若 A 有表达式 $A=M-N$，$M^{-1}\geqslant O$，$N\geqslant O$，则 $\rho(M^{-1}N)<1$.

(Q)：A 非奇异，且 $A^{\mathrm{T}}\boldsymbol{x}\leqslant\boldsymbol{0}$ 与 $\boldsymbol{x}\geqslant\boldsymbol{0}$ 蕴涵 $\boldsymbol{x}=\boldsymbol{0}$.

试证下列的蕴涵关系成立：



2. 设 $A\in\mathbb{Z}^{n\times n}$. 试证：若 A 为非奇异 M-矩阵，则 A 满足题 1 中条件(M)与(N). 反之，若 A 满足条件(D)，(F)，(L)与(P)之一，则 A 为非奇异 M-矩阵.(因此，题 1 中十七个条件(A)～(Q)

当 $A\in Z^{n\times n}$ 时彼此等价，且每一个(因而全部)条件都等价于"A 为 n 阶非奇异 M-矩阵")

3. 证明：对任意 $\boldsymbol{x}=(x_1,\cdots,x_n)^{\mathrm{T}}>\boldsymbol{0}$，$\hat{R}^x=(\hat{R}_1^x,\cdots,\hat{R}_n^x)\in\mathscr{C}_n$ 为 G-函数，且有 $\hat{R}_i^x(A)\leqslant R_i^x(A)$，$\forall A\in\boldsymbol{M}_n(\mathbb{C})$ 与 $\forall i\in\mathscr{N}$，这里 $\hat{R}_i^x(A)$ 由(11)式确定.

4. 证明：若 $A=(a_{ij})\in\boldsymbol{M}_n(\mathbb{C})(n\geqslant2)$ 满足定理 10 条件(4)，则有 $\det A\neq0$.

5. 设 $c_i>0$ 与 $d_i<0$，$i=1,\cdots,n$，试用题 4(取其中的 $f_i(A)=R_i(A)$，$i=1,\cdots,n$)的结果找出下列矩阵 B 为非奇异 M-矩阵的条件：

$$B=\begin{bmatrix} c_1 & d_1 & & & 0 \\ & c_2 & d_2 & & \\ & & \ddots & \ddots & \\ & 0 & & & d_{n-1} \\ d_n & & & & c_n \end{bmatrix}.$$

6. 设 $A=(a_{ij})\in Z^{n\times n}(n\geqslant2)$. 试在不应用推论 9 的限制下，直接证明：条件(18)蕴涵 A 为非奇异 M-矩阵.

8.2 一般 M-矩阵

若用 $\mathscr{K}$ 表示 n 阶非奇异 M-矩阵的集合，用 $\mathscr{K}_0$ 表示 n 阶 M-矩阵的集合，显然 $\mathscr{K}\subset\mathscr{K}_0$，且容易证明：若 $A\in\mathscr{K}_0$，则 $A+\varepsilon I\in\mathscr{K}$，$\forall\varepsilon>0$，因而 $\mathscr{K}$ 的闭包 $\overline{\mathscr{K}}$ 恰是 $\mathscr{K}_0$. 特别地，$\mathscr{K}$ 中的对称矩阵为正定的，$\mathscr{K}_0$ 中的对称矩阵为正半定的.

一般 M-矩阵与非奇异 M-矩阵在应用中几乎同等重要，但或许由于前者(尤其奇异 M-矩阵)研究的难度大，因而它的理论比起后者来至今尚未得以充分地发展.

8.2.1 一般 M-矩阵的特征

我们先来讨论 $\boldsymbol{M}_n(\mathbb{R})$ 中主子式皆为非负实数的矩阵集合的特征. 记此集合为 $\boldsymbol{P}_0$，显然，$\boldsymbol{P}\subset\boldsymbol{P}_0$.

定理 1 设 $A=(a_{ij})\in\boldsymbol{M}_n(\mathbb{R})$. 则下列诸命题是彼此等价的：

(1) $A\in\boldsymbol{P}_0$.

(2) 若 D 为 n 阶正对角矩阵，则 $A+D\in\boldsymbol{P}$.

(3) 对每个非零的实向量 $\boldsymbol{x}=(x_1,\cdots,x_n)^{\mathrm{T}}$，存在 $k\in\mathscr{N}$ 使得 $x_k\neq0$ 且 $x_ky_k\geqslant0$，这里，$\boldsymbol{y}=(y_1,\cdots,y_n)^{\mathrm{T}}=A\boldsymbol{x}$.

(4) 对每个非零的实向量 $\boldsymbol{x}$，存在非负对角矩阵 D_x 使得 $(\boldsymbol{x},D_x\boldsymbol{x})>0$ 与 $(A\boldsymbol{x},D_x\boldsymbol{x})\geqslant0$.

(5) A 的任意主子阵的每一个实特征值非负.

(6) $A+\varepsilon I\in\boldsymbol{P}$，$\forall\varepsilon>0$.

(7) 对每一个只具有 0,1 或 -1 的对角元素的非零对角矩阵 S，存在 $\boldsymbol{y}\neq\boldsymbol{0}$ 使

得 $Sy\geqslant\mathbf{0}$，$S^2y=y$ 与 $SAy\geqslant\mathbf{0}$.

本定理的证明留给读者(见习题 1).

假如在定理 1(4)中令 $D_x=I$，在(7)中令 $S=I$，我们可分别得到如下两个推论.

推论 2　若 $A\in\boldsymbol{M}_n(\mathbb{R})$使得 $A+A^{\mathrm{T}}$ 为对称正半定矩阵，则 $A\in\boldsymbol{P}_0$.

推论 3　若 $A\in\boldsymbol{P}_0$ 则存在 $\boldsymbol{x}\geqslant\mathbf{0},\boldsymbol{x}\neq\mathbf{0}$ 使得 $A\boldsymbol{x}\geqslant\mathbf{0}$.

下面重要定理给出一般 M-矩阵的一些特征.

定理 4　设 $A=(a_{ij})\in Z^{n\times n}$. 则下列诸条件彼此等价：

(1) $A\in\mathscr{K}_0$.

(2) $A+\varepsilon I\in\mathscr{K}$, $\forall\,\varepsilon>0$.

(3) A 的任意主子阵的每一个实特征值非负.

(4) A 的所有主子式非负.

(5) 对每个 $k=1,2,\cdots,n$，A 的所有 k 阶主子式之和为非负实数.

(6) A 的每个实特征值为非负的.

(7) A 的每一个特征值有非负的实部(非负稳定性).

证明　(1)⇒(2)：由 M-矩阵定义即得.

(2)⇒(3)：设$\varnothing\neq\mathscr{M}\subset\mathscr{N}$，且 $\lambda\in\sigma(A[\mathscr{M}])$. 假若 λ 为负实数，按(2)，$B=A-\lambda I\in\mathscr{K}$，因而 $B[\mathscr{M}]$的所有实特征值为正数. 但是，$0\in\sigma(B[\mathscr{M}])=\sigma(A[\mathscr{M}]-\lambda I[\mathscr{M}])$，推出矛盾. 因此，(3)成立.

(3)⇒(4)：这是显然的，因为 A 的主子式等于对应主子阵所有特征值的乘积，而非实特征值之积为正数.

(4)⇒(5)：显然成立.

(5)⇒(6)：类似于 8.1.2 小节定理 3(7)⇒(8)的推证方法，应用 8.1.2 中公式(2)，这时(5)保证此公式中 $c_k\geqslant 0$，$k=1,\cdots,n$. 如果(6)不成立，则 A 有实的负特征值λ_0，即 $\lambda_0<0$ 与 $\det(A-\lambda_0 I)=0$. 将 λ_0 代入 8.1.2 小节公式(2)，其左边为零，而右边为非负数相加之和，且第一项$(-\lambda_0)^n>0$，这是不可能的.

(6)⇒(1)：用类似于 8.1.2 小节定理 3 中(8)⇒(9)的证明方法即得.

(7)⇔(2)：应用 8.1.2 小节定理 3 条件(18)与(9)的等价性即可.　□

由定理 4 可直接得到如下几个推论.

推论 5　设 $A\in\mathscr{K}_0$. 则 A 为非奇异 M-矩阵当且仅当 A 为非奇异的.

证明　设 $A\in\mathscr{K}$，由 8.1.2 小节定理 3 中(9)与(11)的等价性立即推出 A 为非奇异的. 反之，若 $A\in\mathscr{K}_0$ 且 A 非奇异，则按定理 4，A 的每一个实特征值非负，且 $0\notin\sigma(A)$，因而 A 的每一个实特征值必须为正数. 因此，$A\in\mathscr{K}$(见 8.1.2 小节定理 3(8)).　□

推论 6　设 $A\in\boldsymbol{M}_n(\mathbb{R})$. 则 $A\in\boldsymbol{P}_0\cap Z^{n\times n}$ 当且仅当 $A\in\mathscr{K}_0$.

推论 7 设 $A\in\mathbb{Z}^{n\times n}$ 为对称矩阵. 则 $A\in\mathscr{K}_0$ 当且仅当 A 为对称正半定矩阵.

推论 8 设 $A\in\mathbb{Z}^{n\times n}$. 则 $A\in\mathscr{K}_0$ 当且仅当 $A^{\mathrm{T}}\in\mathscr{K}_0$.

推论 9 若 $A\in\mathscr{K}_0$, $B\in\mathbb{Z}^{n\times n}$, 且 $B\geqslant A$, 则 $B\in\mathscr{K}_0$.

证明 两次引用定理 4 条件(2)即可. □

推论 10 若 $A\in\mathbb{Z}^{n\times n}$, 且存在正向量 $\boldsymbol{x}>\boldsymbol{0}$ 使得 $A\boldsymbol{x}\geqslant\boldsymbol{0}$, 则 $A\in\mathscr{K}_0$. 但反之不然.

证明 正命题由定理 4(2)与 8.1.2 小节定理 3(2)立即推出. 逆命题不真, 如取 $A=\begin{bmatrix}0 & -1\\ 0 & 1\end{bmatrix}$, 则 $\sigma(A)=\{0,1\}$, 因而 $A\in\mathbb{Z}^{2\times 2}$ 为 M-矩阵, 但对任意 $\boldsymbol{x}=(x_1,x_2)^{\mathrm{T}}$, $A\boldsymbol{x}\geqslant\boldsymbol{0}$ 蕴涵 $x_2x_1\leqslant 0$. □

由推论 6 与推论 3 看出, 若 $A\in\mathscr{K}_0$, 则有 $\boldsymbol{x}\geqslant\boldsymbol{0}$, $\boldsymbol{x}\neq\boldsymbol{0}$ 使得 $A\boldsymbol{x}\geqslant\boldsymbol{0}$. 但是, 若 $A\in\mathscr{K}_0$ 不可约, 则存在 $\boldsymbol{x}>\boldsymbol{0}$ 使得 $A\boldsymbol{x}\geqslant\boldsymbol{0}$. 这是下面定理的简单推论.

定理 11 设 $A\in\mathscr{K}_0$ 为不可约奇异矩阵. 则

(1) $\operatorname{rank}A=n-1$.

(2) 存在正向量 $\boldsymbol{x}>\boldsymbol{0}$ 使得 $A\boldsymbol{x}=\boldsymbol{0}$.

(3) A 的所有真正主子阵为非奇异 M-矩阵, 特别地有 $a_{ii}>0(1\leqslant i\leqslant n)$.

(4) A 是**几乎单调的**, 即对任意 $\boldsymbol{x}\in\mathbb{R}^n$, $A\boldsymbol{x}\geqslant\boldsymbol{0}$ 蕴涵 $A\boldsymbol{x}=\boldsymbol{0}$.

证明 (1) 设 $A=sI-B$, $s>0$, $B\geqslant O$. 则 $B\in\mathscr{N}_n$ 不可约, 因而由 Perron-Frobenius 定理, $\rho(B)>0$ 为 B 的单重特征值. 又因为 A 奇异, 所以 $s-\rho(B)=0$ 为 A 的单重特征值, 于是, $\operatorname{rank}A=n-1$.

(2) 取 $\boldsymbol{x}$ 为 B 的 Perron 向量, 即有 $\boldsymbol{x}>\boldsymbol{0}$ 与 $B\boldsymbol{x}=\rho(B)\boldsymbol{x}$. 这时, $A\boldsymbol{x}=s\boldsymbol{x}-B\boldsymbol{x}=\rho(B)\boldsymbol{x}-B\boldsymbol{x}=\boldsymbol{0}$.

(3) 在我们假定下, $n\geqslant 2$. 设 $A[\mu]$ 为 A 的某真正主子阵, 这里 μ 为 $\mathscr{N}=\{1,2,\cdots,n\}$ 的真正子集. 则从 $A=sI-B$, $s>0$, $B\geqslant O$ 与 A 的不可约性(因而 B 非负不可约)得出, $A[\mu]=sI[\mu]-B[\mu]$ 是非奇异的 M-矩阵, 这是因为按第七章 7.1.1 小节推论 8, $s\geqslant\rho(B)>\rho(B[\mu])$.

(4) 设对某 $\boldsymbol{x}\in\mathbb{R}^n$, $A\boldsymbol{x}\geqslant\boldsymbol{0}$. 由于 A^{T} 也是不可约的奇异 M-矩阵, 故按(2), 有 $\boldsymbol{y}>\boldsymbol{0}$ 满足 $A^{\mathrm{T}}\boldsymbol{y}=\boldsymbol{0}$ 因而 $\boldsymbol{y}^{\mathrm{T}}A=\boldsymbol{0}^{\mathrm{T}}$. 假若 $A\boldsymbol{x}\neq\boldsymbol{0}$, 则 $\boldsymbol{y}^{\mathrm{T}}A\boldsymbol{x}\neq\boldsymbol{0}$, 此与 $\boldsymbol{y}^{\mathrm{T}}A=\boldsymbol{0}^{\mathrm{T}}$ 矛盾. □

不可约奇异 M-矩阵除了具有上述定理罗列的四个性质以外, 它还具有三角分解性质. 但一般地说, 可约的奇异 M-矩阵不具有三角分解性质. 例如

$$A=\begin{bmatrix}0 & 0 & -1\\ -1 & 0 & -1\\ 0 & 0 & 0\end{bmatrix}\tag{1}$$

没有三角分解, 即不存在下与上三角形矩阵 L 与 U, L 为非奇异 M-矩阵, U 为 M-矩阵, 使得 $A=LU$(见习题 5).

定理 12　设 $A\in\mathscr{K}_0$ 为不可约奇异矩阵. 则存在下三角形非奇异 M-矩阵 L 与上三角形 M-矩阵 U 使得 $A=LU$.

证明　根据定理 11(3),A 的每一个真正主子阵皆为非奇异 M-矩阵. 因此 A 可以写成分块形式

$$A=\begin{bmatrix}A_1 & \boldsymbol{a}\\ \boldsymbol{b}^{\mathrm{T}} & a_m\end{bmatrix},$$

其中,A_1 为 $n-1$ 阶非奇异 M-矩阵,$\boldsymbol{a}\leqslant\boldsymbol{0}$,$\boldsymbol{b}\leqslant\boldsymbol{0}$ 为$\mathbb{R}^{n-1}$中向量,且 $a_m=\boldsymbol{b}^{\mathrm{T}}A_1^{-1}\boldsymbol{a}$,因为在 A 内 A_1 的 Schur 余量$[A/A_1]=a_m-\boldsymbol{b}^{\mathrm{T}}\boldsymbol{A}_1^{-1}\boldsymbol{a}=\det A/\det A_1=0$. 按 8.1.2 小节定理 3(14),存在 $n-1$ 阶的上三角形非奇异 M-矩阵 U_1 与下三角形非奇异 M-矩阵 L_1 使得 $A_1=L_1U_1$. 令

$$L=\begin{bmatrix}L_1 & \boldsymbol{0}\\ \boldsymbol{b}^{\mathrm{T}}U_1^{-1} & 1\end{bmatrix}\quad 与\quad U=\begin{bmatrix}U_1 & L_1^{-1}\boldsymbol{a}\\ \boldsymbol{0}^{\mathrm{T}} & 0\end{bmatrix}.$$

则由于$U_1^{-1}\geqslant O$ 与 $\boldsymbol{b}\leqslant\boldsymbol{0}$,$L_1^{-1}\geqslant O$ 与 $\boldsymbol{a}\leqslant\boldsymbol{0}$,我们有 $\boldsymbol{b}^{\mathrm{T}}U_1^{-1}\leqslant\boldsymbol{0}^{\mathrm{T}}$ 与 $L_1^{-1}\boldsymbol{a}\leqslant\boldsymbol{0}$,因而$L\in\mathscr{K}$与 $U\in\mathscr{K}_0$. 容易看出,$A=LU$. □

由(1)式提供的反例知道,可约奇异 M-矩阵没有定理 12 意义下的 LU 分解,但是,我们可以证明上述定理的一种如下推广.

定理 13　设 $A\in\mathscr{K}_0$. 则存在 n 阶置换矩阵 P,下三角形非奇异 M-矩阵 L 与上三角形 M-矩阵 U 使得 $PAP^{\mathrm{T}}=LU$.

证明　显然地,只要考虑 $A\neq O$ 为可约与奇异的情形. 现令 P 为置换矩阵,使得

$$PAP^{\mathrm{T}}=\begin{bmatrix}A_1 & B\\ O & A_2\end{bmatrix},$$

其中,A_1 为不可约的(或一阶零矩阵). 若 A_2 也是不可约的(或一阶零矩阵),则由定理 12 或 8.1.2 小节定理 3(14),存在下三角形非奇异 M-矩阵 L_1 与 L_2,上三角形 M-矩阵 U_1 与 U_2 使得 $A_1=L_1U_1$ 与 $A_2=L_2U_2$. 此时,因为 $L_1^{-1}B\leqslant O$,我们有

$$L=\begin{bmatrix}L_1 & O\\ O & L_2\end{bmatrix}\in\mathscr{K}\quad 与\quad U=\begin{bmatrix}U_1 & L_1^{-1}B\\ O & U_2\end{bmatrix}\in\mathscr{K}_0$$

满足 $LU=PAP^{\mathrm{T}}$. 若 A_2 还是可约的,则我们可对 A 的可约法式中不可约块的个数应用数学归纳法来证明之. □

我们指出,若定理 13 中 L 改为下三角形 M-矩阵,U 改为上三角形非奇异 M-矩阵,类似结论仍然成立.

对于(1)式中矩阵 A(它是可约奇异 M-矩阵),存在三阶置换矩阵 P

$$P=\begin{bmatrix}0 & 1 & 0\\ 1 & 0 & 0\\ 0 & 0 & 1\end{bmatrix},$$

使得 PAP^{T} 有定理 12 意义下的 LU 分解：

$$PAP^{\mathrm{T}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & -1 & -1 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix}. \tag{2}$$

在本小节最后部分，我们讨论具有规则或弱规则分裂的矩阵的特征以及它们与广义逆正矩阵、M-矩阵之间的关系(关于广义左逆正矩阵的特征可见第五章 5.2.3 小节定理 9).

定理 14　设 $A\in \boldsymbol{M}_n(\mathbb{R})$，并记 $\mathcal{T}_Q=\bigcap_{m=0}^{\infty}\mathrm{Im}Q^m$，$\forall Q\in \boldsymbol{M}_n(\mathbb{R})$. 则下列两个条件彼此等价： □

(1) A 有满足 $\mathcal{T}_{M^{-1}A}=\mathcal{T}_A$ 的规则分裂 $A=M-N$，且迭代矩阵的谱半径不超过 1，亦即：A 有分裂 $A=M-N$ 满足 $M^{-1}\geqslant O, N\geqslant O, \mathcal{T}_{M^{-1}A}=\mathcal{T}_A$，且 $\rho(M^{-1}N)\leqslant 1$.

(2) 存在一个逆正矩阵 B 满足 $B\geqslant A$ 与一个 M-矩阵 $C=I-T, T\geqslant 0$，使得 $\mathcal{T}_C=\mathcal{T}_A, A=BC$.

证明　(1)⇒(2)：设 $A=M-N, M^{-1}\geqslant O, N\geqslant O, \mathcal{T}_{M^{-1}A}=\mathcal{T}_A$，且 $\rho(M^{-1}N)\leqslant 1$. 取 $B=M$，则 $B^{-1}\geqslant O$ 且 $B=A+N\geqslant A$. 又取 $T=M^{-1}N$，则 $T\geqslant O, \rho(T)\leqslant 1$，因而 $C=I-T\in\mathcal{K}_0, \mathcal{T}_C=\mathcal{T}_A$，且 $A=BC$，因为 $C=M^{-1}A$.

(2)⇒(1)：设 $A=BC$，其中 $B\geqslant A, B^{-1}\geqslant O$ 且 $C=I-T, T\geqslant O, \rho(T)\leqslant 1$ 与 $\mathcal{T}_C=\mathcal{T}_A=\bigcap_{m=0}^{\infty}\mathrm{Im}A^m$. 则 $A=B-BT$ 满足 $B^{-1}\geqslant O$ 与 $BT=B-A\geqslant O$. 现令 $M=B$ 与 $N=BT$. 则 $A=M-N$ 为 A 的规则分裂，它满足 $\mathcal{T}_A=\mathcal{T}_C=\mathcal{T}_{M^{-1}A}$ 且 $\rho(M^{-1}N)=\rho(T)\leqslant 1$. □

对于具有弱规则分裂的矩阵(其定义见习题 8.1.3 题 1 中条件(O)或定理 15(1))，我们有下列类似于定理 14 的结果.

定理 15　设 $A\in \boldsymbol{M}_n(\mathbb{R})$，并记 $\mathcal{T}_Q=\bigcap_{m=0}^{\infty}\mathrm{Im}Q^m$，$\forall Q\in \boldsymbol{M}_n(\mathbb{R})$. 则下列两个条件彼此等价：

(1) A 有满足 $\mathcal{T}_{M^{-1}A}=\mathcal{T}_A$ 的弱规则分裂 $A=M-N$，且迭代矩阵的谱半径不超过 1，亦即：A 有分裂 $A=M-N$ 满足 $M^{-1}\geqslant O, M^{-1}N\geqslant O, \mathcal{T}_{M^{-1}A}=\mathcal{T}_A$，且 $\rho(M^{-1}N)\leqslant 1$.

(2) 存在一个逆正矩阵 B 与 M-矩阵 C，使得 $\mathcal{T}_C=\mathcal{T}_A, A=BC$.

证明　(1)⇒(2)：设 $A=M-N, M^{-1}\geqslant O, M^{-1}N\geqslant O, \mathcal{T}_{M^{-1}A}=\mathcal{T}_A$，且 $\rho(M^{-1}N)\leqslant 1$. 令 $B=M$ 与 $C=I-T, T=M^{-1}N$. 则 $B^{-1}\geqslant O$，C 为 M-矩阵，$\mathcal{T}_C=\mathcal{T}_A$，且 $A=BC$，因为 $C=M^{-1}A$.

(2)⇒(1)：假定 $C=sI-\hat{T}, s>0, T=\hat{T}/s\geqslant O, \rho(T)\leqslant 1$，取 $M=sB, N=sBT$，容易看出，$M^{-1}\geqslant O, M^{-1}N=T\geqslant O, M^{-1}A=I-T=C/s$，因此，$A=M-N$ 为满足

(1)中条件的 A 的弱规则分裂. □

若限定 $A\in\mathbb{Z}^{n\times n}$,则前述两组性质连同广义左逆正性都是 M-矩阵的特征. 确切地说我们有

定理 16 设 $A\in\mathbb{Z}^{n\times n}$. 则下列条件彼此等价:

(1) $A\in\mathscr{K}_0$.

(2) A 满足定理 14 中条件(1)或(2).

(3) A 满足定理 15 中条件(1)或(2).

(4) A 为广义左逆正的矩阵.

证明 (1)⇒(2):设 $A\in\mathscr{K}_0, A=sI-B, s>0, B\geqslant O, s\geqslant\rho(B)$. 取 $M=sI$ 与 $N=B$. 则 $M^{-1}\geqslant O, N\geqslant O, \mathscr{T}_{M^{-1}A}=\mathscr{T}_A=\bigcap_{m=0}^{\infty}\mathrm{Im}A^m$, 且 $\rho(M^{-1}N)=\rho(B)/s\leqslant 1$. 因此,(2)成立.

(2)⇒(3):显然成立.

(3)⇒(4):根据第五章 5.2.3 小节定理 9,(4)等价于以下性质:A 在 $\mathscr{T}_A=\bigcap_{m=0}^{\infty}\mathrm{Im}A^m$ 上单调,也就是说,$A\boldsymbol{x}\geqslant\boldsymbol{0}$ 与 $\boldsymbol{x}\in\mathscr{T}_A$ 蕴涵 $\boldsymbol{x}\geqslant\boldsymbol{0}$. 现设 $A\boldsymbol{x}\geqslant\boldsymbol{0}$ 与 $\boldsymbol{x}\in\mathscr{T}_A$. 由(3),$A=BC, B^{-1}\geqslant O, C\in\mathscr{K}_0, \mathscr{T}_C=\mathscr{T}_A$. 因而 $C\boldsymbol{x}=B^{-1}A\boldsymbol{x}\geqslant\boldsymbol{0}, \boldsymbol{x}\in\mathscr{T}_C=\mathscr{T}_A$. 令 $C=sI-\hat{T}, s>0, \hat{T}\geqslant O, s\geqslant\rho(\hat{T})$,则 $C=s(I-T)$,这里 $T=\hat{T}/s\geqslant O, \rho(T)\leqslant 1$. 这时由后面的引理 17 知道,$C$ 的 Drazin 逆 $C^{\mathrm{D}}=s^{-1}(I-T)^{\mathrm{D}}$ 在 $\mathscr{T}_C=\mathrm{Im}(I-T)^k(k=\mathrm{index}(I-T))$上非负. 因此,$\boldsymbol{x}=C^{\mathrm{D}}C\boldsymbol{x}\geqslant\boldsymbol{0}$,因为 C 的 Drazin 逆 C^{D} 满足 $C^{\mathrm{D}}C\boldsymbol{y}=\boldsymbol{y}, \forall\boldsymbol{y}\in\mathscr{T}_C$,且 $C\boldsymbol{x}\in\mathrm{Im}C^{k+1}=\mathrm{Im}C^k=\mathscr{T}_C$.

(4)⇒(1):设(4)成立. 根据第五章 5.2.3 小节定理 9,(4)等价于如下性质:A 的每一个广义左逆在 $\mathscr{T}_A$ 上非负. 现设 $A=sI-B, s>0, B\geqslant O$,且令 $T=B/s$. 因 $A^{\mathrm{D}}=s^{-1}(I-T)^{\mathrm{D}}$ 为 A 的一个广义左逆,所以 A^{D}(因而$(I-T)^{\mathrm{D}}$)在 $\mathscr{T}_A$ 上非负. 再应用后面的引理 17 便得 $\rho(T)\leqslant 1$,即 $s\geqslant\rho(B)$. 因此,$A\in\mathscr{K}_0$. □

余下我们来证明前一定理推证中已经应用过的一个引理.

引理 17 设 $T\in\mathscr{N}_n$. 则 $\rho(T)\leqslant 1$ 当且仅当$(I-T)^{\mathrm{D}}$ 在 $\mathrm{Im}(I-T)^k$ 上非负,这里 $k=\mathrm{index}(I-T)$. 并且,若 $\rho(T)\leqslant 1, 0<\alpha<1$ 与 $T_\alpha=(1-\alpha)I+\alpha T$,则

$$(I-T)^{\mathrm{D}}=\alpha\sum_{j=0}^{\infty}T_\alpha^jE, \tag{3}$$

这里,$E=(I-T)(I-T)^{\mathrm{D}}$.

证明 设$\rho(T)\leqslant 1$. 若$\rho(T)<1$,则 $I-T\in\mathscr{K}$,因而有$(I-T)^{\mathrm{D}}=(I-T)^{-1}\geqslant O$,且$(I-T)^{-1}=\sum_{j=0}^{\infty}T^j=\alpha\sum_{j=0}^{\infty}T_\alpha^j$,因为 $I-T=\alpha^{-1}(I-T_\alpha), 0<\alpha<1$. 因此只要考虑 $\rho(T)=1$ 的情形. 这时,$T_\alpha\geqslant O, 0<\alpha<1$,且容易看出,$\sigma(T_\alpha)=(1-\alpha)+\alpha\sigma(T)$ 包含在单位闭圆内且它在单位圆周上只有元素 1. 由于 $I-T_\alpha=\alpha(I-T)$,

故有$(I-T)^{\mathrm{D}}=\alpha(I-T_{\alpha})^{\mathrm{D}}$与$E=(I-T)(I-T)^{\mathrm{D}}=(I-T_{\alpha})(I-T_{\alpha})^{\mathrm{D}}$. 考虑$T_{\alpha}$与$E$的Jordan正规形式,并注意到这两个正规形式中相似变换矩阵可取为相同的事实,我们有$\rho(T_{\alpha}E)<1, T_{\alpha}E=ET_{\alpha}$,并且,

$$(I-T_{\alpha})^{\mathrm{D}}=(I-T_{\alpha}E)^{-1}+E-I.$$

根据Neumann引理,

$$(I-T_{\alpha}E)^{-1}=I+\sum_{j=1}^{\infty}T_{\alpha}^{j}E,$$

因为E为幂等矩阵. 因此,

$$(I-T)^{\mathrm{D}}=\alpha(I-T_{\alpha})^{\mathrm{D}}=\alpha\sum_{j=0}^{\infty}T_{\alpha}^{j}E,$$

此即(3)式. 现设$\boldsymbol{x}\geqslant\boldsymbol{0}, \boldsymbol{x}\in\operatorname{Im}(I-T)^{k}$,这里,$k=\operatorname{index}(I-T)$. 由于$E$为沿着$\operatorname{Ker}(I-T)^{k}$到$\operatorname{Im}(I-T)^{k}$上的投影,故有$E\boldsymbol{x}=\boldsymbol{x}$. 因为$T_{\alpha}\geqslant O(0<\alpha<1)$,所以从(3)式得出,

$$(I-T)^{\mathrm{D}}\boldsymbol{x}=\alpha\sum_{j=0}^{\infty}T_{\alpha}^{j}E\boldsymbol{x}=\alpha\sum_{j=0}^{\infty}T_{\alpha}^{j}\boldsymbol{x}\geqslant\boldsymbol{0}.$$

此表明$(I-T)^{\mathrm{D}}$在$\operatorname{Im}(I-T)^{k}$上非负.

反之,设$(I-T)^{\mathrm{D}}$在$\operatorname{Im}(I-T)^{k}$上非负,$k=\operatorname{index}(I-T)$,且$T\boldsymbol{x}=\rho(T)\boldsymbol{x}, \boldsymbol{x}\geqslant\boldsymbol{0}, \boldsymbol{x}\neq\boldsymbol{0}$. 当$\rho(T)\neq1$时,

$$(I-T)\boldsymbol{x}=(1-\rho(T))\boldsymbol{x},$$

因而

$$(I-T)^{j}\boldsymbol{x}=(1-\rho(T))^{j}\boldsymbol{x}, \forall j\geqslant0.$$

上式蕴涵$\boldsymbol{x}\in\bigcap_{m=0}^{\infty}\operatorname{Im}(I-T)^{m}=\operatorname{Im}(I-T)^{k}$. 此时,

$$\boldsymbol{x}=(I-T)^{\mathrm{D}}(I-T)\boldsymbol{x}=(1-\rho(T))(I-T)^{\mathrm{D}}\boldsymbol{x}\geqslant\boldsymbol{0}.$$

但由于$\boldsymbol{x}\neq\boldsymbol{0}$与$(I-T)^{\mathrm{D}}\boldsymbol{x}\geqslant\boldsymbol{0}$,我们有$1-\rho(T)>0$亦即$\rho(T)<1$. 因此$\rho(T)\leqslant1$.

□

习题 8.2.1

1. 证明定理1,并证明$\boldsymbol{P}_0$为$\boldsymbol{P}$的闭包.

2. 试举出反例说明定理1中命题(7)不能改换为如下命题:“对每一个符号矩阵S,存在向量$\boldsymbol{x}\geqslant\boldsymbol{0}, \boldsymbol{x}\neq\boldsymbol{0}$使得$SAS\boldsymbol{x}\geqslant\boldsymbol{0}$”.

3. 设$A=(a_{ij})\in\boldsymbol{M}_n(\mathbb{R})$,用下列固定的定母表示关于$A$的某些条件:

(A) A的所有主子式非负.

(B) A的所有k阶主子式之和非负,$k=1,\cdots,n$.

(C) A的每一个实特征值非负.

(D) A的每一个非零特征值的实部为正数.

(E) A 为非负稳定的,即 A 的每一个特征值的实部非负.

试证:

(1) (A)⇒(B)⇒(C).

(2) (D)⇒(E)⇒(C).

4. 当限定 $A=(a_{ij})\in Z^{n\times n}$ 时,试证:题 3 中条件(A)～(E)分别等价于 $A\in\mathscr{K}_0$.

5. 证明:(1)式中三阶矩阵 A 为奇异可约的M-矩阵,它不存在定理12意义下的 LU 三角分解.

6. 证明:若 $A,B\in\mathscr{K}_0$ 且 $AB\in Z^{n\times n}$,则 $AB\in\mathscr{K}_0$.

7. 设 $A\in\boldsymbol{M}_n(\mathbb{R})$. 则下列诸命题彼此等价:

(1) $A\in\mathscr{K}_0$.

(2) 对每一个正对角矩阵 D,$A+D$ 为逆正的.

(3) 对每一个正实数 α,$A+\alpha I$ 为逆正的.

8. 设 $A,B\in\mathscr{K}_0$,且"∘"表示矩阵的 Hadamard 积. 试证:

(1) $A\circ B$ 的比较矩阵 $\mathscr{M}(A\circ B)\in\mathscr{K}_0$.

(2) 若 $B\in\mathscr{K}$,则 $\mathscr{M}(A\circ B^{-1})\in\mathscr{K}_0$.

8.2.2 带有"性质 c"的 M-矩阵

本小节研究 $\mathscr{K}_0$ 中一个重要的子类,它包含 $\mathscr{K}$ 与不可约奇异 M-矩阵等. 这个 $\mathscr{K}_0$ 的子类具有类似于非奇异 M-矩阵的一些基本性质,因而有广泛的应用价值,尤其在奇异线性代数方程组迭代法求解中有着深入的应用.

定义 1 $A=(a_{ij})\in\mathscr{K}_0$ 称为具有**"性质 c"的矩阵**,假如 A 可以表达为 $A=sI-B$, $s>0$, $B\geqslant O$,使得矩阵 $T=B/s$ 的幂序列有极限,即 $\lim\limits_{j\to\infty}T^j$ 存在.

回忆一下(见第二章 2.2.2 小节定理 7),对 $T\in\boldsymbol{M}_n(\mathbb{R})$,则 $\lim\limits_{j\to\infty}T^j$ 存在当且仅当

(1) $\rho(T)\leqslant 1$;

(2) 若 $\rho(T)=1$,则 $\operatorname{index}(I-T)\leqslant 1$,即对应 T 特征值 1 的初等因子皆为线性的;

(3) 若 $\rho(T)=1$,则 $\lambda\in\sigma(T)$ 与 $|\lambda|=1$ 蕴涵 $\lambda=1$.

显然,所有非奇异 M-矩阵都具有"性质 c". 然而,不是所有的奇异 M-矩阵都分享这个性质的. 例如矩阵

$$A=\begin{bmatrix}0 & -1\\0 & 0\end{bmatrix}\tag{1}$$

为奇异 M-矩阵,若 $A=sI-B$, $s>0$, $B\geqslant O$,则 $T=B/s$ 有形式

$$T=\begin{bmatrix}1 & s^{-1}\\0 & 1\end{bmatrix}.$$

因为

$$T^j=\begin{bmatrix}1 & js^{-1}\\0 & 1\end{bmatrix},\qquad j=1,2,\cdots,$$

所以$\lim_{j\to\infty}T^j$不存在. 同时还要注意,若$A\in\mathscr{K}_0$且$A=sI-B,s>\max_i a_{ii},B\geqslant O$,则按Perron-Frobenius定理,如果$\rho(T)=1,\lambda\in\sigma(T)$与$|\lambda|=1$必定蕴涵$\lambda=1$,即$\lim_{j\to\infty}T^j$存在的等价条件中(3)自动成立(但(1)或(2)不一定成立). 事实上,此时,$b_{ii}=s-a_{ii}>0,\forall i$,于是,按第七章7.1.2小节推论7,$T$的法式中每个对角块为不可约素矩阵,因而当$\rho(T)=1$时,按Perron-Frobenius定理,$\lambda\in\sigma(T)$与$|\lambda|=1$蕴涵$\lambda=1$.

我们强调指出,对$A\in\mathscr{K}_0$,按某种分解$A=sI-B,s>0,B\geqslant O,T=B/s$的幂序列不收敛,但按另一种类似的分解,$T$的幂序列可能收敛. 例如考虑$M$-矩阵

$$A=\begin{bmatrix}1 & -1\\ -1 & 1\end{bmatrix}. \tag{2}$$

取$s=1$,则$\lim_{j\to\infty}T^j$不存在;若取$s>1$,则$\lim_{j\to\infty}T^j$便存在,因而按定义1,(2)式中A为具有“性质c”的M-矩阵.

现在考虑具有“性质c”的M-矩阵的特征.

定理2 设$A=(a_{ij})\in\mathscr{K}_0$. 则$A$为具有“性质$c$”的$M$-矩阵当且仅当$\operatorname{index}A\leqslant 1$(即$A$非奇异或$A$奇异但$\operatorname{rank}A^2=\operatorname{rank}A$),或者等价地,$A$的群逆$A^{\#}$存在.

证明 只要设A为奇异的M-矩阵,且$A=sI-B,s>\max_i a_{ii},B\geqslant O,s=\rho(B)$. 令$T=B/s$,此时按前面说明,$\lim_{j\to\infty}T^j$存在的等价条件中(3)自动成立,且存在非奇异矩阵$Q$使得$QTQ^{-1}$为$T$的Jordan正规形式:

$$QTQ^{-1}=\begin{bmatrix}J & O\\ O & H\end{bmatrix}, \tag{3}$$

其中,$\rho(H)<1,\rho(J)=\rho(T)=1$. 于是,

$$Q(I-T)Q^{-1}=\begin{bmatrix}I-J & O\\ O & I-H\end{bmatrix}, \tag{4}$$

且式中$I-H$非奇异. 根据第二章2.2.2小节定理7,这时,$\lim_{j\to\infty}T^j$存在当且仅当$I-J=O$,但此条件又等价于$\operatorname{rank}(I-T)=\operatorname{rank}(I-T)^2$. 因为$A=s(I-T)$,这意味着$\operatorname{rank}A=\operatorname{rank}A^2$,这表明:$\operatorname{index}A=1$. □

具有“性质c”M-矩阵另一个类似于非奇异M-矩阵的特征如下(见8.1.2小节定理3(17)与第七章习题7.2.2题4与题5).

定理3 设$A\in Z^{n\times n}$. 则A为具有“性质c”的M-矩阵,当且仅当存在一个对称正定矩阵W使得$AW+WA^{\mathrm{T}}$为对称正半定矩阵.

证明 设存在对称正定矩阵W使得$AW+WA^{\mathrm{T}}$正半定. 根据第四章习题4.2.2题9,$\operatorname{rank}W=n\leqslant\pi(A)+\tau(A)\leqslant n$,其中$\tau(A)$表示$A$在虚轴上特征值的初等因子数,因而$\pi(A)+\tau(A)=\pi(A)+\nu(A)+\delta(A)=n$,于是有$\tau(A)=\nu(A)+\delta(A)$. 由于$\tau(A)\leqslant\delta(A)$,我们有$\nu(A)=0$与$\tau(A)=\delta(A)$. 按8.2.1小节定理4

(7),$A\in\mathcal{K}_0$. 现设 $A=sI-B, s>\max\limits_i a_{ii}, B\geqslant O, s\geqslant\rho(B)$,且 $T=B/s$,则 $\rho(T)\leqslant 1$,且由 $\tau(A)=\delta(A)$ 知道,对应于 A 的零特征值的初等因子皆为线性的,因而对应于 T 的特征值 1 的初等因子也皆为线性的. 因此,$\lim\limits_{j\to\infty}T^j$ 存在,从而 A 为具有"性质 c"的 M-矩阵.

反之,设 $A\in\mathbb{Z}^{n\times n}$ 为具有"性质 c"的 M-矩阵,且 $A=sI-B, s>\max\limits_i a_{ii}, B\geqslant O$. 令 $T=B/s$,则有 $\rho(T)\leqslant 1$. 类似于定理 2 的证明,这时,(3)与(4)两式也成立. 由于 $\rho(H)<1$,故 $I-H$ 为非奇异 M-矩阵,因而按 8.1.2 小节定理 3(18)与第四章 4.2.1 小节定理 6,存在对称正定矩阵 $\widetilde{W}$ 使得 $(I-H)\widetilde{W}+\widetilde{W}(I-H)^{\mathrm{T}}$ 为对称正定的. 置 $W_1\in\boldsymbol{M}_n(\mathbb{R})$ 为

$$W_1=\begin{bmatrix} I & O \\ O & \widetilde{W}\end{bmatrix},$$

与 $\widetilde{A}=Q(I-T)Q^{-1}$(见(4)式). 则 $\widetilde{A}=\mathrm{diag}[O,I-H]$,并且,

$$\begin{aligned}\widetilde{A}W_1+W_1\widetilde{A}^{\mathrm{T}}&=\begin{bmatrix} O & O \\ O & I-H\end{bmatrix}\begin{bmatrix} I & O \\ O & \widetilde{W}\end{bmatrix}+\begin{bmatrix} I & O \\ O & \widetilde{W}\end{bmatrix}\begin{bmatrix} O & O \\ O & I-H^{\mathrm{T}}\end{bmatrix}\\ &=\begin{bmatrix} O & O \\ O & (I-H)\widetilde{W}+\widetilde{W}(I-H)^{\mathrm{T}}\end{bmatrix}\equiv C\end{aligned}$$

为对称正半定的. 现令 $W=Q^{-1}W_1(Q^{-1})^{\mathrm{T}}$,则显然,$W$ 也是对称正定的,且 $A=s(I-T)$ 满足 $AW+WA^{\mathrm{T}}=Q^{-1}(sC)(Q^{-1})^{\mathrm{T}}$,其中右边矩阵对称正半定. □

由前面知道(1)式中二阶奇异 M-矩阵不具有"性质 c". 这个结论也可以由定理 3 证实. 事实上,若存在二阶对称正定矩阵 $W=(w_{ij})$ 使得 $AW+WA^{\mathrm{T}}$ 为正半定的. 则从

$$AW+WA^{\mathrm{T}}=\begin{bmatrix} -w_{12}-w_{21} & -w_{22} \\ -w_{22} & 0\end{bmatrix}$$

知道,右边矩阵的行列式等于 $-w_{22}^2<0$,推出矛盾.

下一个结果告诉我们,若 $A\in\mathbb{Z}^{n\times n}$ 为广义对角占优矩阵,则 A 为具有"性质 c"的 M-矩阵.

定理 4 若 $A\in\mathbb{Z}^{n\times n}$,且存在 $\boldsymbol{x}>\boldsymbol{0}$ 使得 $A\boldsymbol{x}\geqslant\boldsymbol{0}$,则 A 为具有"性质 c"的 M-矩阵.

本定理证明留给读者(见习题 1). 我们指出,定理 4 的逆命题不真. 例如,

$$A=\begin{bmatrix} 0 & -1 \\ 0 & 1\end{bmatrix}$$

为奇异 M-矩阵,它具有"性质 c",因为 $A=A^2$. 但不存在 $\boldsymbol{x}>\boldsymbol{0}$ 使得 $A\boldsymbol{x}\geqslant\boldsymbol{0}$.

应用定理 4 与 8.2.1 小节定理 11(2),立即得出如下结果.

推论 5 不可约奇异 M-矩阵具有"性质 c".

在本小节最后部分，我们讨论具有"性质 c"的 M-矩阵与矩阵的广义逆正性(见第五章 5.2.3 小节)之间的关系.

定理 6　设 $A\in\mathbb{Z}^{n\times n}$. 则下列诸命题彼此等价：

(1) A 为具有"性质 c"的 M-矩阵.

(2) $\operatorname{index}A\leqslant 1$ 且存在逆正矩阵 $M\geqslant A$ 与 M-矩阵 $B=I-T$，$T\geqslant O$，使得 $\lim\limits_{j\to\infty}T^j$ 存在，$\operatorname{Im}B=\operatorname{Im}A$ 与 $A=MB$.

(3) $\operatorname{index}A\leqslant 1$ 且存在逆正矩阵 M 与具有"性质 c"的 M-矩阵 B，使得 $\operatorname{Im}B=\operatorname{Im}A$ 与 $A=MB$.

(4) $\operatorname{index}A\leqslant 1$ 且 A 有一个非负的$\{1\}$-逆 B，使得 $\operatorname{Im}(BA)=\operatorname{Im}A$，也就是说，存在 $B\geqslant O$ 使得 $A=ABA$ 与 $\operatorname{Im}(BA)=\operatorname{Im}A$.

(5) $\operatorname{index}A\leqslant 1$ 且 A 有一个非负的$\{1\}$-逆 B，使得 $\operatorname{Im}(BA)\dot{+}\operatorname{Ker}A=\mathbb{R}^n$.

(6) 对 A 的每一个规则分裂 $A=M-N$，都有 $\rho(M^{-1}N)\leqslant 1$ 与 $\operatorname{index}(I-M^{-1}N)\leqslant 1$.

证明　(1)⇒(2)：设 $A\in\mathbb{Z}^{n\times n}$ 为具有"性质 c"的 M-矩阵. 则 A 有规则分裂 $A=sI-C$，$s>0$，$C\geqslant O$ 且 $T=C/s$ 的幂列收敛. 取 $M=sI$，$B=I-T$，便知(2)成立.

(2)⇒(3)：显然.

(3)⇒(4)：设(3)成立，$A=MB$，其中 $M^{-1}\geqslant O$，B 为具有"性质 c"的 M-矩阵，$\operatorname{Im}B=\operatorname{Im}A$. 令 $B=sI-C$，$s>\max\limits_i b_{ii}$，$C\geqslant O$，$s\geqslant\rho(C)$，$T=C/s$ 使得 $\lim\limits_{j\to\infty}T^j$ 存在. 于是根据定理 2 与 8.2.1 小节引理 17，$(I-T)^{\mathrm{D}}$ 在 $\operatorname{Im}(I-T)=\operatorname{Im}B=\operatorname{Im}A$ 上非负. 由于 $\operatorname{index}B\leqslant 1$，故 $\operatorname{index}(I-T)\leqslant 1$，因而$(I-T)^{\mathrm{D}}$ 为 $I-T$ 的一个$\{1\}$-逆. 现设 $A\boldsymbol{x}\geqslant\mathbf{0}$ 与 $\boldsymbol{x}\in\operatorname{Im}A$. 这时，$B\boldsymbol{x}=M^{-1}MB\boldsymbol{x}=M^{-1}A\boldsymbol{x}\geqslant\mathbf{0}$，因而 $\boldsymbol{x}=B^{\mathrm{D}}B\boldsymbol{x}\geqslant\mathbf{0}$(因为 $B^{\mathrm{D}}=s^{-1}(I-T)^{\mathrm{D}}$ 在 $\operatorname{Im}B$ 上非负)，这表示 A 在 $\operatorname{Im}A$ 上单调. 但 $\operatorname{index}A\leqslant 1$ 且 A 在 $\operatorname{Im}A$ 上单调等价于(4)(见习题 2)，因此，(4)成立.

(4)⇒(5)：显然成立，因为按第五章 5.2.2 小节定理 2，$\operatorname{index}A\leqslant 1$ 蕴涵 $\operatorname{Im}A\dot{+}\operatorname{Ker}A=\mathbb{R}^n$.

(5)⇒(6)：设(5)成立. 先证对 A 的任意给定的规则分裂 $A=M-N$，$\sum\limits_{i=0}^{\infty}(M^{-1}N)^iM^{-1}\boldsymbol{x}$ 对所有 $\boldsymbol{x}\geqslant\mathbf{0}$ 且 $\boldsymbol{x}\in\operatorname{Im}A$ 收敛. 令 $S_p=\sum\limits_{i=0}^{p-1}(M^{-1}N)^i$，$p=1,2,\cdots$. 则 $S_pM^{-1}A=I-(M^{-1}N)^p$，$p=1,2,\cdots$，因而 $B-S_pM^{-1}AB=(M^{-1}N)^pB$，$p=1,2,\cdots$. 但是按第五章 5.1.1 小节定理 6(8)，$ABy=\boldsymbol{y}$，$\forall\,\boldsymbol{y}\in\operatorname{Im}A$，我们有

$$\mathbf{0}\leqslant B\boldsymbol{x}-S_pM^{-1}\boldsymbol{x}=(M^{-1}N)^pB\boldsymbol{x},\qquad\forall\,\boldsymbol{x}\geqslant\mathbf{0}\text{ 且 }\boldsymbol{x}\in\operatorname{Im}A.$$

这表明向量列$\{S_pM^{-1}\boldsymbol{x}\}_{p=1}^{\infty}$ 有上界 $B\boldsymbol{x}$，且由于此列单调不减，故它收敛. 于是，$\sum\limits_{i=0}^{\infty}(M^{-1}N)^iM^{-1}\boldsymbol{x}$ 收敛，$\forall\,\boldsymbol{x}\geqslant\mathbf{0}$ 且 $\boldsymbol{x}\in\operatorname{Im}A$. 接着证明：$\rho(M^{-1}N)\leqslant 1$ 与 $\operatorname{index}(I-$

$M^{-1}N)\leqslant 1$.

假若 $\rho\equiv\rho(M^{-1}N)>1$,且 $\boldsymbol{y}\neq\boldsymbol{0},\boldsymbol{y}\geqslant\boldsymbol{0}$ 满足 $M^{-1}N\boldsymbol{y}=\rho\boldsymbol{y}$,则 $M\boldsymbol{y}=\rho^{-1}N\boldsymbol{y}\geqslant\boldsymbol{0}$,并且,$M\boldsymbol{y}=N\boldsymbol{y}+A\boldsymbol{y}$ 蕴涵 $M\boldsymbol{y}=(1-\rho)^{-1}A\boldsymbol{y}\in\operatorname{Im}A$. 这时用刚才证明结果有 $\boldsymbol{0}\leqslant\sum_{i=0}^{\infty}\rho^{i}\boldsymbol{y}=\sum_{i=0}^{\infty}(M^{-1}N)^{i}M^{-1}(M\boldsymbol{y})<\infty$,但此与 $\rho>1$ 矛盾. 因此,$\rho=\rho(M^{-1}N)\leqslant 1$. 若 $\rho<1$,则显然有 $\operatorname{index}(L-M^{-1}N)\leqslant 1$. 现设 $\rho=1$,因而 $1\in\sigma(M^{-1}N)$. 假若有 $\operatorname{index}(I-M^{-1}N)>1$,按第七章 7.2.2 小节定义 6 及其后说明,$M^{-1}N$ 有特征向量 $\boldsymbol{y}\geqslant\boldsymbol{0}$ 对应于 1 使得对某个二级广义特征向量 $\boldsymbol{x}\neq\boldsymbol{0}$,$(M^{-1}N-I)\boldsymbol{x}=\boldsymbol{y}$. 于是,$M\boldsymbol{y}=N\boldsymbol{y}\geqslant\boldsymbol{0}$ 与 $M\boldsymbol{y}=-A\boldsymbol{x}\in\operatorname{Im}A$. 因此,

$$-\lim_{p\to\infty}(I-(M^{-1}N)^{p})\boldsymbol{x}=\lim_{p\to\infty}S_{p}\boldsymbol{y}=\lim_{p\to\infty}\sum_{i=0}^{p-1}(M^{-1}N)^{i}M^{-1}(M\boldsymbol{y})<\infty,$$

这表明 $\{(M^{-1}N)^{p}\boldsymbol{x}\}_{p=1}^{\infty}$ 收敛. 但从 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 关系,我们有 $(M^{-1}N)^{p}\boldsymbol{x}=\boldsymbol{x}+p\boldsymbol{y}$,$p=1,2,\cdots$,因而 $\{(M^{-1}N)^{p}\boldsymbol{x}\}_{p=1}^{\infty}$ 发散,推出矛盾.

(6)⇒(1):设(6)对 $A\in Z^{n\times n}$ 成立,因 $A=sI-B(s>0,B\geqslant O)$ 是 A 的一种规则分裂,所以 $\rho(B/s)\leqslant 1$ 与 $\operatorname{index}(I-B/s)\leqslant 1$,因而按定义与定理 2,$A$ 为具有"性质 c"的 M-矩阵. □

习题 8.2.2

1. 证明定理 4.(提示:用第七章习题 7.2.2 题 4,5)

2. 设 $A\in\boldsymbol{M}_n(\mathbb{R})$,$\mathscr{S}\subset\mathbb{R}^n$. 试证下列诸命题彼此等价:

(1) A 有一个非负$\{1\}$-逆 B,$\operatorname{Im}(BA)=\mathscr{S}$.

(2) A 有一个$\{1\}$-逆 B,$\operatorname{Im}(BA)=\mathscr{S}$,且 B 在 $\operatorname{Im}A$ 上非负.

(3) A 有一个$\{1,2\}$-逆 C,$\operatorname{Im}C=\mathscr{S}$,且 C 在 $\operatorname{Im}A$ 上非负.

(4) A 的满足条件 $\operatorname{Im}C=\mathscr{S}$ 的每一个$\{1,2\}$-逆 C 在 $\operatorname{Im}A$ 上非负.

(5) $\mathscr{S}\dot{+}\operatorname{Ker}A=\mathbb{R}^n$ 且 A 在 $\mathscr{S}$ 上单调,即 $A\boldsymbol{x}\geqslant\boldsymbol{0}$ 与 $\boldsymbol{x}\in\mathscr{S}$ 蕴涵 $\boldsymbol{x}\geqslant\boldsymbol{0}$.

3. 设 $A\in\boldsymbol{M}_n(\mathbb{R})$,且 $A=M-N$ 为 A 的规则分裂:$M^{-1}\geqslant O,N\geqslant O$. 试证下列两个命题等价:

(1) $\rho(M^{-1}N)\leqslant 1$ 与 $I-M^{-1}N$ 的群逆 $(I-M^{-1}N)^{\#}$ 存在.

(2) A 满足题 2 的某个条件,其中 $\mathscr{S}\subset\mathbb{R}^n$.

4. 设 $A\in Z^{n\times n}$ 为 M-矩阵. 若 $\operatorname{index}A\leqslant k$,$k=0,1,\cdots$,则称 A 为 **M-k-矩阵**. 所有 n 阶 M-k-矩阵的集合记为 $\mathfrak{M}_k$. 显然,$\mathfrak{M}_0=\mathscr{K}$,$\mathfrak{M}_1$ 为具有"性质 c"的 n 阶 M-矩阵集合. 试证:若 $A=sI-B$,$s>\max_i a_{ii}$,$B\geqslant O$,则下列诸命题彼此等价:

(1) $A\in\mathfrak{M}_k$.

(2) 序列$\{s^{-j}B^{j}\}$有低于 k 次的多项式展开,即存在次数$<k$ 的多项式 $\psi(j)$ 使得 $\lim_{j\to\infty}(s^{-j}B^{j}-\psi(j))=O$(当 $k=0$ 时,$\lim_{j\to\infty}s^{-j}B^{j}=O$).

(3) $\lim_{j\to\infty}j^{-k+1}s^{-j}B^{j}$ 存在.

(4) 序列$\{j^{-k+1}s^{-j}B^{j}\}$有界.

(5) $\lim\limits_{j\to\infty} j^{-k}s^{-j}B^j = O$.

5. 设 $T\in\mathcal{N}_n$ 满足 $\rho(T)=1$,且 $A=I-T$, $T_\alpha=(1-\alpha)I+\alpha T$, $\alpha\in(0,1)$. 试证下列诸命题彼此等价:

(1) A 为具有"性质 c"的 M-矩阵.

(2) T 的任意两个基类不彼此相通.

(3) 对某个 $\alpha\in(0,1)$, $\lim\limits_{j\to\infty} T_\alpha^j$ 存在.

(4) 对每一个 $\alpha\in(0,1)$, $\lim\limits_{j\to\infty} T_\alpha^j$ 存在.

6. 设 $A\in Z^{n\times n}$. 证明: $A\in\mathfrak{M}_k$(见题 4)当且仅当 $\nu=\mathrm{index}A\leqslant k$ 且 A^{D} 在 $\mathrm{Im}A^{\nu}=\mathrm{Im}A^k$ 上非负.

8.2.3　M-矩阵与有限齐次 Markov 链

在第七章 7.3.1 小节中,我们已经介绍了有限齐次 Markov 链的某些术语与有关结果. 应用本章前面介绍的 M-矩阵理论可以得到齐次 Markov 链的更深入的分析结果. 在下面有限齐次 Markov 链将简称为 Markov 链.

定理 1　若 T 为某 Markov 链的转移矩阵,则 $A=I-T$ 为具有"性质 c"的 M-矩阵,因而 A 的群逆 $A^{\#}$ 存在 .

证明　由于 $T\geqslant O$ 与 $\rho(T)=1$, $A=I-T$ 显然为奇异的 M-矩阵. 余下只要证 $\mathrm{rank}A=\mathrm{rank}A^2$. 若 T 为非负不可约矩阵,则按第七章 7.3.1 小节定理 6(1),以 T 为转移矩阵的 Markov 链为遍历的,这时, $1\in\sigma(T)$ 为单重特征值,因而 $0\in\sigma(A)$ 也是单重的,且 A 的 Jordan 正规形式有如下表示:

$$J_A=\begin{bmatrix}0 & \mathbf{0}^{\mathrm{T}}\\ \mathbf{0} & K\end{bmatrix}, \tag{1}$$

式中 K 非奇异. 因此, $\mathrm{rank}A=\mathrm{rank}A^2$,按 8.2.2 小节定理 2,这表明 $A=I-T$ 为具有"性质 c"的 M-矩阵.

若 T 为非负可约矩阵,不失普遍性,设该链的所有遍历类排列在瞬态类之前,即 T 有形式

$$T=\begin{bmatrix}D_1 & O & \cdots & O & O\\ O & D_2 & \cdots & O & O\\ \vdots & & \ddots & \vdots & \vdots\\ & & & D_r & \\ B_1 & B_2 & \cdots & B_r & C\end{bmatrix}, \tag{2}$$

其中 $D_1,\cdots,D_r$ 不可约且谱半径均为 1, C 对应着所有瞬态类且 $\rho(C)<1$(见习题 1). 将本证明的前一部分结果用于 $I-D_j$ 便知道 $I-D_j$ 为具有"性质 c"的 M-矩阵. 显然, $I-C$ 为非奇异 M-矩阵. 因此, $A=I-T$ 为具有"性质 c"的 M-矩阵.　□

定理 2　设 T 为某 Markov 链的转移矩阵, $A=I-T$,且 $L=I-AA^{\#}$. 则当且

仅当 $\delta(T)=\max\{|\lambda|:\lambda\in\sigma(T),\lambda\neq\rho(T)\}<1$,即 1 是 T 在单位圆周上的唯一特征值时,$\lim\limits_{k\to\infty}T^k$ 存在. 此时,$\lim\limits_{k\to\infty}T^k=L$.

证明 按定理 1,$A=I-T$ 为具有"性质 c"的 M-矩阵,因而 $A^{\#}$ 存在. 这时,$\lim\limits_{k\to\infty}T^k$存在当且仅当 $\delta(T)<1$(见第二章 2.2.2 小节定理 7). 但若$\lim\limits_{k\to\infty}T^k$ 存在,则有非奇异矩阵 P 使得

$$T=P\begin{bmatrix}I & O\\ O & K\end{bmatrix}P^{-1},\tag{3}$$

其中 $\rho(K)<1$,因而

$$A=I-T=P\begin{bmatrix}O & O\\ O & I-K\end{bmatrix}P^{-1},$$

$$(I-T)^{\#}=P\begin{bmatrix}O & O\\ O & (I-K)^{-1}\end{bmatrix}P^{-1}=A^{\#}.\tag{4}$$

容易验证,$\lim\limits_{k\to\infty}T^k=P\begin{bmatrix}I & O\\ O & O\end{bmatrix}P^{-1}=I-AA^{\#}=L.$ □

按 Markov 链的术语我们有如下推论.

推论 3 一个遍历的 Markov 链的转移矩阵 T 的幂列有极限当且仅当此链是规则的. 一个吸收链的转移矩阵 T 总使$\lim\limits_{k\to\infty}T^k$ 存在. □

证明 第一部分结果由第七章 7.3.1 小节定理 6(1)与习题 7.3.1 题 4 即得. 若一个随机矩阵 T 为某吸收链的转移矩阵,则这样的链的每个遍历类只由单个吸收态组成,经过行列同步变换,T 可写成(2)式的形式,其中 $r=1$ 与 $D_1=I$,且 $\rho(C)<1$. 于是,$\lim\limits_{k\to\infty}T^k$ 存在. □

应用定理 2 与推论 3 以及 T 的表达式(2),我们有

定理 4 设 T 为一个 Markov 链的转移矩阵. 则$\lim\limits_{k\to\infty}T^k$ 存在当且仅当存在置换矩阵 P 使得

$$PTP^{\mathrm{T}}=\mathrm{diag}[T_{11},T_{22},\cdots,T_{mm}],\tag{5}$$

其中每一个 T_{ii}有形如(2)式的结构,且(2)式中每一个 D_i 为素矩阵. 此时,

$$\lim_{k\to\infty}T^k=P^{\mathrm{T}}\mathrm{diag}[L_{11},L_{22},\cdots,L_{mm}]P,\tag{6}$$

其中,$L_{ii}=I-(I-T_{ii})(I-T_{ii})^{\#},i=1,\cdots,m.$

下面定理指出,应用简单的特征值平移变换每一个转移矩阵 T 可以变为幂列收敛的矩阵 T_α.

定理 5 设 T 为某 Markov 链的转移矩阵,$A=I-T$ 与 $L=I-AA^{\#}$. 则矩阵 $T_\alpha=(1-\alpha)I+\alpha T(0<\alpha<1)$的幂列收敛,且有

$$\lim_{k\to\infty}T_\alpha^k=L.\tag{7}$$

证明　对 $\alpha\in(0,1)$，T_α 为随机矩阵，因为 $T_\alpha\geqslant O$ 且对 $e=(1,\cdots,1)^{\mathrm{T}}$，$T_\alpha e=(1-\alpha)e+\alpha Te=e$. 此外，$\sigma(T_\alpha)$ 在单位圆周上只有元素 1(见 8.2.1 小节引理 17 的证明)，按定理 2，$\lim\limits_{k\to\infty}T_\alpha^k=I-(I-T_\alpha)(I-T_\alpha)^{\#}=I-AA^{\#}$，因为 $I-T_\alpha=\alpha(I-T)=\alpha A$. □

作为上述定理的推论，我们可以导出遍历链的唯一的正平稳概率分布向量的表达式(见第七章 7.3.1 小节推论 9).

定理 6　设 T 为某遍历链的转移矩阵，$A=I-T$ 与 $L=I-AA^{\#}$. 则

$$L=e\pi=\begin{bmatrix}\pi\\ \pi\\ \vdots\\ \pi\end{bmatrix},\tag{8}$$

其中，$e=(1,\cdots,1)^{\mathrm{T}}$，$\pi$ 为该链唯一的正平稳概率分布(行)向量.

证明　按(7)式，L 为(行)随机矩阵. 此外，$LA=(I-AA^{\#})A=A-A=O$，即 $LT=L$. 因此，$LT^k=L$，$\forall k\geqslant 1$. 根据第七章 7.3.1 小节推论 9，L 的每一个行向量必等于 π. □

从下面几个结果可以看出矩阵 $A^{\#}$ 与 $L=I-AA^{\#}$ 在 Markov 链分析中的重要作用，这里，$A=I-T$.

定理 7　设某 Markov 链有状态集合 $\{s_1,\cdots,s_n\}$，它的转移矩阵为 T. 则 s_i 为瞬态当且仅当 $L=I-AA^{\#}$ 的第 i 列元素全为零. 等价地，s_i 为遍历态当且仅当 $Le_i\neq\mathbf{0}$，这里 e_i 为 $\mathbb{R}^n$ 的第 i 个单位坐标向量.

证明　不失一般性，设转移矩阵 T 有形式(2)，将它改写为更紧凑形式

$$T=\begin{bmatrix}D & O\\ B & C\end{bmatrix},\tag{9}$$

其中，$D=\mathrm{diag}[D_1,\cdots,D_r]$，诸 D_i 对应着遍历类，而所有瞬态对应着矩阵 C. 因此，$A=I-T$ 有形式

$$A=\begin{bmatrix}A_{11} & O\\ A_{21} & A_{22}\end{bmatrix},\tag{10}$$

其中，$A_{11}=I-D$，$A_{21}=-B$，$A_{22}=I-C$. 此时，$\mathrm{index}A_{11}=1$，$\rho(C)<1$(见习题 1)，因而 $I-C$ 为非奇异 M-矩阵. 于是，A 的群逆有形式

$$A^{\#}=\begin{bmatrix}A_{11}^{\#} & O\\ * & A_{22}^{-1}\end{bmatrix},\tag{11}$$

从而 $L=I-AA^{\#}$ 有形式

$$L=\begin{bmatrix}L_{11} & O\\ * & O\end{bmatrix},\qquad L_{11}=I-A_{11}A_{11}^{\#}.\tag{12}$$

根据定理 6,L_{11}的每一列至少有一非零元素,因为 $A_{11}=I-D=I-\mathrm{diag}[D_1,\cdots,D_r]$,而 D_i 为某个遍历链的转移矩阵,$i=1,\cdots,r$. □

定理 8　设某 Markov 链有状态集合$\{s_1,\cdots,s_n\}$,它的转移矩阵为 T,又设 $A=I-T$ 与 $L=I-AA^{\#}$. 若状态 s_i 与 s_j 属于同一遍历类,则 L 的第 i 行与第 j 行相同,即有 $\boldsymbol{e}_i^{\mathrm{T}}L=\boldsymbol{e}_j^{\mathrm{T}}L$. □

证明　如前所述,可设 T 有形式(9),因而 L 有形式(12). 这时由于遍历类对应于 D 的对角块 D_i,若 s_i 与 s_j 都在同一遍历类内,则按定理 6 与前一定理的证明,L_{11}的第 i 与第 j 行相同. □

定理 9　设某 Markov 链有状态集合$\{s_1,\cdots,s_n\}$,其转移矩阵为 T,又设 $A=I-T$与 $L=I-AA^{\#}$. 记$[s_k]$为该链包含某个固定状态 s_k 的遍历类,$\mathscr{I}_k$ 为$[s_k]$中状态的指标集合. 则对于 $L=(l_{ij})$与 $1\leqslant i\leqslant n$,

$$\sum_{j\in\mathscr{I}_k} l_{ij} = \text{该过程从 } s_i \text{ 最终被吸收到}[\,s_k]\text{ 的概率}. \tag{13}$$

证明　不妨设 T 有形式(2)或(9),且令

$$\widetilde{T}=\begin{bmatrix} I & O \\ B & C\end{bmatrix}\in \mathscr{N}_n.$$

则以 $\widetilde{T}$ 为转移矩阵的 Markov 链显然为吸收链. 同时,过程从状态 $\tilde{s}_i$ 最终吸收到对应于$(\widetilde{T})_{kk}=1$ 的状态 $\tilde{s}_k$ 的概率为

$$(\lim_{m\to\infty}\widetilde{T}^m)_{ik} = (\widetilde{L})_{ik},$$

其中$(B)_{ik}$ 表示矩阵B 的(i,k) 位置上的元素,且 $\widetilde{L}=I-(I-\widetilde{T})(I-\widetilde{T})^{\#}$. 由此推出,在原先 Markov 链中,$s_i$ 最终被吸收到$[s_k]$ 的概率等于 $\sum\limits_{j\in\mathscr{I}_k}\tilde{l}_{ij}$,这里,$\widetilde{L}=(\tilde{l}_{ij})$. 再应用 L 与 $\widetilde{L}$ 的关系便得(13) 式. □

回顾一下第七章 7.3.1 小节例 5 中的 Markov 链,在那里状态 s_1 与 s_2 为吸收态,$\{s_1\}$ 与$\{s_2\}$ 为遍历类,而状态 s_3,s_4 与 s_5 分别为瞬态,$\{s_3,s_4,s_5\}$ 组成该链唯一的瞬态类. 由计算结果$\lim\limits_{m\to\infty}T^m=L$ 看出,从状态 s_3 最终变为$[s_1]=\{s_1\}$ 的概率为 $\sum\limits_{j\in\mathscr{I}_1}l_{3j}=l_{31}=7/15$,即甲由初始状态(1元钱)最终变为输光状态的概率是7/15,类似地,甲最终为胜者的概率是 $l_{32}=8/15$,等等.

当我们假定 Markov 链至少有一个瞬态时,$A^{\#}$的元素也给出链的有用信息.

定理 10　设某 Markov 链有状态集合$\{s_1,\cdots,s_n\}$,其转移矩阵为 T,又设 $A=I-T$,且 $A^{\#}=(a_{ij}^{\#})$. 若 s_i 与 s_j 为瞬态,则

$$a_{ij}^{\#} = \text{过程初始在 } s_i \text{ 后变为 } s_j \text{ 的次数的期望值}. \tag{14}$$

证明　不妨设 T 有形式(2)或(9). 这时,s_i 与 s_j 和矩阵 C 对应,而 $\rho(C)<1$(见习题 1),因而$(I-C)^{-1}=\sum\limits_{m=0}^{\infty}C^m=A_{22}^{-1}$(见(11)式). 但是,如(9)式中矩阵 D 为 d

阶的，则

$$\left(\sum_{m=0}^{k-1} C^m\right)_{i-d,j-d} = \left(\sum_{m=0}^{k-1} T^m\right)_{ij}$$

恰是过程初始在状态 s_i，k 步内在状态 s_j 的次数的期望值，于是，过程初始在 s_i 后变为 s_j 的次数的期望值应为

$$\lim_{k\to\infty}\left(\sum_{m=0}^{k-1} C^m\right)_{i-d,j-d} = ((I-C)^{-1})_{i-d,j-d} = (A_{22}^{-1})_{i-d,j-d} = a_{ij}^{\#}. \qquad \square$$

对于第七章 7.3.1 小节例 5 中的 Markov 链来说，

$$(I-C)^{-1} = \frac{1}{5}\begin{bmatrix} 7 & 6 & 4 \\ 3 & 9 & 6 \\ 1 & 3 & 7 \end{bmatrix},$$

$$A^{\#} = \left[\begin{array}{cc:ccc} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \hdashline & & \frac{7}{5} & \frac{6}{5} & \frac{4}{5} \\ & * & \frac{3}{5} & \frac{9}{5} & \frac{6}{5} \\ & & \frac{1}{5} & \frac{3}{5} & \frac{7}{5} \end{array}\right],$$

因此，该过程最初在 s_3（瞬态）以后在 s_4（瞬态）的次数的期望值等于 $a_{34}^{\#}=\frac{6}{5}$，等等.

习题 8.2.3

1. 设 $T\in\mathcal{N}_n$ 为随机矩阵，它有形式(2). 试证：$\rho(C)<1$.
2. 考虑有如下转移矩阵 T 的 Markov 链：

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 0 & 1 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{2}{3} & 0 & 0 & \frac{1}{6} & \frac{1}{6} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

试对每个链确定遍历类与瞬态类，给出形式(2)的结构. 并且，以它们为例验证一下本小节内的有关结论.

3. 证明：若 T 与 T^{-1} 分别为某个 Markov 链的转移矩阵，则 T 必与置换矩阵.
4. 考虑具有如下转移矩阵的 Markov 链：

$$T = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

(1) 证明此链为规则的.

(2) 求解 $\pi T=\pi(\pi_1+\pi_2+\pi_3=1)$以确定该链的平稳概率分布向量.

(3) 确定 $L=\lim\limits_{k\to\infty}T^k$.

5. 设 T 为某遍历 Markov 链的转移矩阵,$A=I-T$. 证明:$A^\#=A^\dagger$ 当且仅当 T 为双随机矩阵.

8.3　数理经济学中的投入-产出模型分析

本节研究 1973 年度诺贝尔经济学奖得主 W. W. Leontief 的静态投入-产出模型,它是最简单的生产模型. 应用 *M*-矩阵理论为主要数学工具,可大为简化此模型的分析.

8.3.1　引言与开式 Leontief 模型

Leontief 投入-产出分析处理如下特殊的经济问题:在一种特定的经济状态中,它的每一个产业为了正好满足社会上各经济部门对其产品的总需求,应具有怎样的产出水平?

在投入-产出开式模型中,对于生产条件规定得特别严格,而对消费行为却无限制——实际上将消费作为任意给定的非负最终需求向量. 在进行静态分析时,我们不对一个过程进行时间方面的分析,只是对两种不同的均衡进行比较,其中一种均衡是由另一种均衡在某些参数变动下产生出来的. 确切地说,在 Leontief 方法里,一个经济体的生产活动被分散成 n 个产业部门(然而不必要按微观意义分成一些单个商行),并且分析各部门间的产品交易. 其基本假定如下:

(1) 不存在联合生产,n 个产业部门中每一个只生产单一产品. 用数学语言来说,这意味着部门与产品之间存在着一一对应关系,因而部门与产品这两个概念可相互代用. 今后,用 i 表示生产第 i 种产品的部门或第 i 种产品本身.

(2) 每个部门只有一种生产方法,生产意味着一定量的几种产品变换成一定数额的单一种产品. 而且,这个投入-产出变换的模式是稳定的,这表明部门 j 为生产一个单位的第 j 种产品需要 t_{ij} 单位的 i 产品作为投入,$i=1,\cdots,n$,并且 j 产品的 λ 单位的产出需要第 i 种产品 λt_{ij} 单位. 诸量 t_{ij} 叫作**投入系数**,它通常假定是不变的. 按经济学术语,投入比例为常数,且这个常数反过来粗略地估计投入比例.

(3) 不存在生产滞后;不存在资本品;模型不考虑反复进入生产过程超出一个生产阶段的商品,而只包括一经过程使用便不复存在的商品.

(4) 不存在对外贸易也不牵连政府的活动.

按照上述基本假定,我们可以给出这种最简单生产模型的数学描述. 用 x_i 表示每固定单位时间内第 i 种产品的**总产出**,$\boldsymbol{x}=(x_1,\cdots,x_n)^{\mathrm{T}}$ 表示**产出向量**. 总产

出 x_i 的一部分作为 n 个部门生产活动所需要的投入而用掉了，即有

$$\sum_{j=1}^{n} t_{ij} x_j$$

单位的 i 产品消耗在生产活动中，余下

$$d_i = x_i - \sum_{j=1}^{n} t_{ij} x_j \tag{1}$$

单位的 i 产品作为净产出. 这个净产出 d_i 正规地叫作第 i 种产品的**最终需求**. 用 $\boldsymbol{d}=(d_1,\cdots,d_n)^{\mathrm{T}}$ 表示**最终需求向量**，它可视为外部部门(居民)的消费向量. 这时方程(1)表明，对产品 i 来说，总需求等于总供给，进而我们得到描述整个经济体的总投入-产出基本平衡方程组

$$(I-T)\boldsymbol{x} = \boldsymbol{d}, \tag{2}$$

这里，$T=(t_{ij})\in \boldsymbol{M}_n(\mathbb{R})$ 叫作**投入矩阵**，而 $A=I-T\in \mathbb{Z}^{n\times n}$ 通常叫作 **Leontief 型矩阵**，或**基本非正矩阵**. 上述模型叫作**开式 Leontief 模型**. 显而易见，这个模型的特性完全取决于 $A=I-T$ 的性质.

我们的问题是要研究经济体的可行性，即对任取外部需求向量 $\boldsymbol{d}\in\mathbb{R}_+^n$，方程组(2)是否都有一个非负解 $\boldsymbol{x}\in\mathbb{R}_+^n$. 显然，当且仅当 $A=I-T$ 为非奇异 M-矩阵时，对任意 $\boldsymbol{d}\in\mathbb{R}_+^n$，方程组(2)有非负解 $\boldsymbol{x}\in\mathbb{R}_+^n$. 事实上，若 $A\in\mathscr{K}$，则有 $A^{-1}\geqslant O$(逆正性)，因而 $\boldsymbol{x}=A^{-1}\boldsymbol{d}\geqslant\boldsymbol{0}$，$\forall\,\boldsymbol{d}\in\mathbb{R}_+^n$，反之，若对任意 $\boldsymbol{d}\in\mathbb{R}_+^n$，方程组(2)有解 $\boldsymbol{x}\in\mathbb{R}_+^n$，则 $A=I-T$ 有逆，因而从 $\boldsymbol{x}=A^{-1}\boldsymbol{d}\geqslant\boldsymbol{0}$，$\forall\,\boldsymbol{d}\in\mathbb{R}_+^n$ 立即推出 $A^{-1}\geqslant O$. 于是，$A\in\mathbb{Z}^{n\times n}$ 为非奇异 M-矩阵(见 8.1.2 小节定理 3(11)与(9)的等价性).

上述经济模型同时有一个相应的价格-价值系统，它给出投入-产出关系的价格与价值的侧面. Leontief 模型的价格理论认为，一种商品的价格等于其单位生产成本加上单位产出的净收入. 现引入下列记号：

p_j：商品 j 的价格；v_j：商品 j 单位产出的增殖价值，即单位产出的净收入.

则显然

$$\sum_{i=1}^{n} t_{ij} p_i, \quad 1\leqslant j\leqslant n$$

为商品 j 的单位生产成本，于是，商品 j 单位产出增殖为

$$v_j = p_j - \sum_{i=1}^{n} t_{ij} p_i, \quad 1\leqslant j\leqslant n. \tag{3}$$

若用 $\boldsymbol{p}=(p_1,\cdots,p_n)^{\mathrm{T}}$ 表示相应的 Leontief 模型的**价格向量**，$\boldsymbol{v}=(v_1,\cdots,v_n)^{\mathrm{T}}$ 表示**增殖向量**，则(3)式可改写为矩阵形式

$$\boldsymbol{p}^{\mathrm{T}} - \boldsymbol{p}^{\mathrm{T}} T = \boldsymbol{v}^{\mathrm{T}}, \tag{4}$$

但由于 $A=I-T$，我们有

$$\boldsymbol{p}^{\mathrm{T}} A = \boldsymbol{v}^{\mathrm{T}}. \tag{5}$$

显然，$\boldsymbol{v}\in\mathbb{R}_+^n$，而且我们兴趣在于考察方程组(4)或(5)的非负价格向量的解 $\boldsymbol{p}$.

下述关系为方程组(2)与(4)之间的桥梁：

$$\sum_{i=1}^{n} v_i x_i = \sum_{j=1}^{n} p_j d_j, \tag{6}$$

因为 $A\boldsymbol{x}=\boldsymbol{d}$ 与方程组(5)蕴涵 $\boldsymbol{p}^{\mathrm{T}}A\boldsymbol{x}=\boldsymbol{p}^{\mathrm{T}}\boldsymbol{d}=\boldsymbol{v}^{\mathrm{T}}\boldsymbol{x}$. 可以把(6)式翻译成经济学的说法："国民收入"等于"国民生产".

原先产出方程组(2)对任意 $\boldsymbol{d}\in\mathbb{R}_+^n$ 有非负产出向量 $\boldsymbol{x}$ 的解意味着模型的**可行性**，而价格方程组(4)或(5)对任意增值向量 $\boldsymbol{v}\in\mathbb{R}_+^n$ 有非负的价格向量 $\boldsymbol{p}$ 的解则意味着模型为**可获利的**(profitable). 由于方程组(5)等价于方程组

$$A^{\mathrm{T}}\boldsymbol{p}=\boldsymbol{v}, \tag{7}$$

故类似于模型可行性的分析，模型为可获利的当且仅当 A^{T}(因而 A)为非奇异 M-矩阵.

因此，我们有如下基本定理，在这里，以上两个概念的对偶性变得十分明显.

定理 1　设开式 Leontief 模型有投入矩阵 $T\in\mathcal{N}_n$，且 $A=I-T$. 则下列命题彼此等价：

(1) 模型为可行的.

(2) 模型为可获利的.

(3) A 为非奇异 M-矩阵.

回忆一下，当 $A\in\mathbb{Z}^{n\times n}$ 为非奇异 M-矩阵时，本章 8.1 节已给出它的许多等价性质. 当前一定理中(3)改换为这些等价性质之一时，定理显然仍成立. 例如，当(3)换为 8.1.2 小节定理 3(6)，即 A 的所有主子式为正数时，便得到经济学文献中关于开式Leontief 模型可行性(或可获利性)的著名的 **Hawkins-Simon 条件**. 但实际上这个条件早在 12 年前已由 Ostrowski 建立.

假如开式 Leontief 模型的投入矩阵 $T=(t_{ij})\in\mathcal{N}_n$ 为不可约的(此可直观地解释为该经济体内任一部门都直接或间接地有赖于其他部门的生产)，则 $A=I-T\in\mathbb{Z}^{n\times n}$ 也不可约，因而按 8.1.2 小节定理 7，我们有 A 为非奇异 M-矩阵等价于：$A^{-1}>O$. 此时，定理 1 有如下推论.

推论 2　设开式 Leontief 模型有不可约的投入矩阵 $T\in\mathcal{N}_n$，且 $A=I-T$. 则下列命题彼此等价：

(1) 方程组(2)对任意 $\boldsymbol{d}\geqslant\boldsymbol{0}(\neq\boldsymbol{0})$ 有正向量的解 $\boldsymbol{x}$.

(2) 方程组(5)对任意 $\boldsymbol{v}\geqslant\boldsymbol{0}(\neq\boldsymbol{0})$ 有正向量的解 $\boldsymbol{p}$.

(3) $A^{-1}>O$.

现在讨论**效应问题**或**相对静态分析**，即一个可行模型的开式需求 $\boldsymbol{d}$ 的变化将对最终产出 $\boldsymbol{x}$ 可能产生什么效果，或者一个可行模型增殖需要 $\boldsymbol{v}$ 的变化将对价格 $\boldsymbol{p}$ 可能产生什么影响. 假如此模型的投入矩阵 T 为可约的，那么显然地，对某些商

品最终需求的增加，将可能不影响其他商品的总产出. 例如，$T=(t_{ij})\in\mathcal{N}_2$ 可约，$A=I-T=(a_{ij})\in Z^{2\times2}$有形式 $A=\begin{bmatrix}a_{11} & a_{12}\\ 0 & a_{22}\end{bmatrix}$. 则方程组

$$\begin{bmatrix}a_{11} & a_{12}\\ 0 & a_{22}\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix}=\begin{bmatrix}d_1\\ d_2\end{bmatrix}$$

中 d_1 的变动不会引起 x_2 的变动，而 d_2 的变动会引起 x_1 与 x_2 的变动. 为此目的先需要如下巧妙的引理及其推论.

引理 3 设 $A=(a_{ij})\in Z^{n\times n}$ 满足严格对角占优条件：$a_{ii}>-\sum\limits_{j\neq i}a_{ij}, i=1,\cdots,n$. 则 A^{-1} 存在且它的元素 $(A^{-1})_{ij}$ 满足

$$(A^{-1})_{ii}>(A^{-1})_{ki},\qquad \forall k\neq i. \tag{8}$$

证明 由于题设条件，故按 8.1.2 小节定理 3 中命题(3)与(9)的等价性可知，A 为非奇异 M-矩阵，因而 $A^{-1}\geqslant O$. 倘若不等式(8)不真，则存在某 i_0 与 k_0，$i_0\neq k_0$，使得 $(A^{-1})_{i_0i_0}\leqslant(A^{-1})_{k_0i_0}$. 这时不妨假定：$(A^{-1})_{k_0i_0}\geqslant(A^{-1})_{li_0}$，$\forall l\in\mathcal{N}=\{1,2,\cdots,n\}$. 从 $AA^{-1}=I_n$ 的事实看出，

$$a_{k_0k_0}(A^{-1})_{k_0i_0}+\sum_{j\neq k_0}a_{k_0j}(A^{-1})_{ji_0}=0,$$

但因为 $A^{-1}\geqslant O$ 与 $(A^{-1})_{k_0i_0}>0$，所以有

$$\begin{aligned}0<a_{k_0k_0}(A^{-1})_{k_0i_0}&=-\sum_{j\neq k_0}a_{k_0j}(A^{-1})_{ji_0}\\ &\leqslant-\sum_{j\neq k_0}a_{k_0j}(A^{-1})_{k_0i_0}<a_{k_0k_0}(A^{-1})_{k_0i_0}.\end{aligned}$$

但这是不可能的. 因此，不等式(8)成立. □

从引理 3 我们可以得到下面的推论.

推论 4 设 $A=(a_{ij})\in Z^{n\times n}$.

(1) 如果 A 为非奇异M-矩阵，且有 $a_{ii}\geqslant-\sum\limits_{j\neq i}a_{ij}, i=1,\cdots,n$，则 A^{-1} 的元素 $(A^{-1})_{ij}$ 满足

$$(A^{-1})_{ii}\geqslant(A^{-1})_{ki},\qquad i,k=1,\cdots,n. \tag{9}$$

(2) 如果 A 不可约且有 $a_{ii}>-\sum\limits_{j\neq i}a_{ij}, i=1,\cdots,n$，则

$$(A^{-1})_{ii}>(A^{-1})_{ki}>0,\qquad \forall k\neq i. \tag{10}$$

证明 (1) 应用引理 3 做连续性推证便得所要的结论.

(2) 按题设这时 A 是不可约非奇异 M-矩阵，因而 $A^{-1}>O$. 于是，不等式(10)便是不等式(8)的直接推论. □

我们指出，引理 3 与推论 4 中关于 $a_{ii}\geqslant-\sum\limits_{j\neq i}a_{ij}, i=1,\cdots,n$ 的假定即使对具

有投入矩阵 T 的可行开式 Leontief 模型也不一定成立，例如，$A = I - T$，其中

$$T = \begin{bmatrix} \frac{3}{7} & \frac{5}{7} \\ \frac{2}{7} & \frac{1}{7} \end{bmatrix}. \tag{11}$$

但是，假如这个条件成立，即 T 的每个行和至多为 1，则我们有如下结果.

定理 5 设 $T \in \mathcal{N}_n$ 为某可行的开式 Leontief 模型的投入矩阵，且 T 的每个行和至多为 1. 则有

(1) 当商品 i 的最终需求单独增加时，各商品的产出不减少.

(2) 此时，商品 i 的产出一定增加，且增加的幅度最大，虽然其他商品的产出也可能按此幅度增加.

证明 如前，令 $A = I - T$，$\boldsymbol{x} = (x_1, \cdots, x_n)^{\mathrm{T}}$ 与 $\boldsymbol{d} = (d_1, \cdots, d_n)^{\mathrm{T}}$ 分别表示产出与最终需求向量. 按定理 1，A 为非奇异 M- 矩阵，按题设，A 满足条件 $a_{ii} \geqslant -\sum_{j \neq i} a_{ij}$，$i = 1, \cdots, n$，于是，可以应用推论 4(1). 现设 $\boldsymbol{d}$ 的第 i 个分量 d_i 有一个正的改变量 δ，其余分量不变. 则由此产生的新需求向量为 $\hat{\boldsymbol{d}} = \boldsymbol{d} + \delta \boldsymbol{e}_i$，这里 $\boldsymbol{e}_i$ 表示 $\mathbb{C}^n$ 的第 i 个单位坐标向量，因而新产出向量 $\hat{\boldsymbol{x}}$ 为

$$\hat{\boldsymbol{x}} = A^{-1} \hat{\boldsymbol{d}} = A^{-1}(\boldsymbol{d} + \delta \boldsymbol{e}_i) = A^{-1}\boldsymbol{d} + \delta A^{-1} \boldsymbol{e}_i,$$

然而由于 $\boldsymbol{x} = A^{-1}\boldsymbol{d}$，故有

$$\hat{\boldsymbol{x}} - \boldsymbol{x} = \delta A^{-1} \boldsymbol{e}_i \text{ 或等价地 } \hat{x}_k - x_k = \delta (A^{-1})_{ki}, \quad 1 \leqslant k \leqslant n. \tag{12}$$

由上式推出，$\hat{x}_i > x_i$，因为 $(A^{-1})_{ii} > 0$（见(9)式），且 $\hat{x}_j \geqslant x_j$，$j = 1, \cdots, n$，因为 $A^{-1} \geqslant O$，即有(1). 类似地，结论(2)也是(9)与(12)式的直接推论. □

当投入矩阵 T 不可约且其所有行和均小于 1 时，应用推论 4(2)，可以得出上述定理的改进结果.

推论 6 设不可约 $T \in \mathcal{N}_n$ 为某开式 Leontief 模型的投入矩阵，且 T 的所有行和均小于 1. 则有：

(1) 当商品 i 的最终需求单独增加时，所有商品的产出将增加.

(2) 此时，商品 i 产出增加幅度大于其他商品的增加幅度.

证明 应用(12)与(10)式即可. □

假如将对应于商品 k 最终需求 d_k 的商品 i 总产出 x_i 的**弹性**定义为

$$e_{ik} = \frac{d_k}{x_i} \cdot \frac{\Delta x_i}{\Delta d_k}, \quad \Delta x_i = \hat{x}_i - x_i, \qquad \Delta d_k = \hat{d}_k - d_k. \tag{13}$$

那么，在推论 6 假设前提下，若 $\Delta d_k > 0$，$\Delta d_j = 0$，$\forall j \neq k$，则 $\Delta x_i / \Delta d_k = (A^{-1})_{ik} > 0$. 并且，下面推论指出，$e_{ik}$ 总小于 1.

推论 7 在推论 6 假设前提下，$e_{ik} < 1$，$\forall i, k$.

证明　令 $\hat{x}_j=\beta_j x_j,\beta_j>1,\forall j;\hat{d}_k=d_j,\forall j\neq k$，且 $\hat{d}_k=\gamma d_k,\gamma>1$. 则有

$$e_{ik}=\frac{d_k(\beta_i-1)x_i}{x_i(\gamma-1)d_k}=\frac{\beta_i-1}{\gamma-1}.$$

对于固定的 k，若有 i 使得 $e_{ik}\geqslant 1$，则有 $\beta_i\geqslant\gamma$. 由方程 $x_k=\sum_{j=1}^{n}t_{kj}x_j+d_k$ 与 $\hat{x}_j,\hat{d}_k$ 定义得出

$$\hat{x}_k/\beta_k=\sum_{j=1}^{n}t_{kj}\hat{x}_j/\beta_j+\hat{d}_k/\gamma. \tag{14}$$

但对新的最终需求向量有

$$\hat{x}_k=\sum_{j=1}^{n}t_{kj}\hat{x}_j+\hat{d}_k. \tag{15}$$

按推论 6，$\beta_k>\beta_j,\forall j\neq k$，因而 $\beta_j^{-1}>\beta_k^{-1},\forall j\neq k$. 再由 $\gamma^{-1}\geqslant\beta_i^{-1}\geqslant\beta_k^{-1}$，从(15)式必须得出

$$\begin{aligned}\hat{x}_k/\beta_k&=\sum_{j=1}^{n}t_{kj}\hat{x}_j/\beta_k+\hat{d}_k/\beta_k\\&<\sum_{j=1}^{n}t_{kj}\hat{x}_j/\beta_j+\hat{d}_k/\gamma,\end{aligned}$$

因为 $T\in\mathcal{N}_n$ 不可约，所以有 $j\neq k$ 使得 $t_{kj}>0$. 此与(14)式抵触. 因此，$e_{ik}<1$. □

前面考察了最终需求变动对于总产出的相对静态效应，由于线性方程组(2)与(4)之间有对偶关系，我们可考虑增殖需要变化后对于价格的相对静态效应问题. 这时除了用 T 的列和替代行和外，增殖与价格之间也具有前一种效应问题的类似关系. 不过，若一个开式 Leontief 模型是可获利的，那么该经济体中没有一个部门经营蚀本，且至少有一个部门经营生利. 对于投入矩阵 $T=(t_{ij})\in\mathcal{N}_n$，这意味着 $\sum_{i=1}^{n}t_{ij}x_j\leqslant x_j,\forall j$，且至少对一个 j，严格不等式成立，因而有

$$\sum_{i=1}^{n}t_{ij}\leqslant 1,\qquad j=1,\cdots,n, \tag{16}$$

且上述不等式中至少有一严格成立. 这个重要事实导出定理 5 的如下对偶结论.

定理 8　设 $T=(t_{ij})\in\mathcal{N}_n$ 为某可行开式 Leontief 模型的投入矩阵. 则有

(1) 当商品 i 的增殖单独增加时，没有一种商品价格下跌.

(2) 此时，商品 i 的价格一定上涨，且上涨幅度最大，虽然其他商品价格也可能按此幅度上涨.

证明　用 $A^{\mathrm{T}}=I-T^{\mathrm{T}}$ 代替定理 5 中的 $A=I-T$，并用事实(16). □

对于 T 为不可约情形，我们也可以给出推论 6 的一个对偶结果.

推论 9　设不可约 $T\in\mathcal{N}_n$ 为某开式 Leontief 模型的投入矩阵，且 T 的所有列和均小于 1. 则有：

(1) 当商品 i 的增殖单独增加时,所有商品价格上涨.

(2) 此时,商品 i 的价格上涨幅度大于其他商品的上涨幅度.

我们指出,定理 5 中关于 T 的行和的假定是必要的,但由于事实(16),定理 8 中关于 T 的每个列和至多为 1 的假定自然成立. 作为例证,考虑以(11)式中二阶矩阵 T 为投入矩阵的开式的 Leontief 模型. 此时 T 的最大行和大于 1,但由于 T 的最大列和小于 1,故该模型为可行的因而也是可获利的. 令 $A=I-T$. 则 A 为非奇异 M-矩阵(不可约的),且

$$A^{-1}=\frac{1}{2}\begin{bmatrix}6 & 5\\ 2 & 4\end{bmatrix}.$$

此模型不满足定理 5. 譬如,若商品 2 的最终需求单独增加一个单位,那么产出向量按 $\Delta x=(5/2,2)^{\mathrm{T}}$ 增加,这表明商品 1 的产出增加幅度大于商品 2 的. 另一方面,应用推论 9 便知,增殖单个分量的增加导致所有价格分量的增加,且前者对应的价格增加最大. 在本例中这可由 A^{-1} 为严格行对角占优的事实来解释.

另一个对于价格的相对静态效应问题与引入税收(增大某些增殖)与补贴(减少某些增殖)有关. 为简单起见,假定对商品 1 征税,而对商品 2 补贴,且税收量与补贴量都等于 $\tau>0$.

定理 10　设 $T\in\mathcal{N}_n$ 为某可行的开式 Leontief 模型的投入矩阵,且 T 的所有行和至多为 1. 若对商品 1 征税 $\tau>0$,对商品 2 补贴 τ,则商品 1 的价格不会下跌,而商品 2 的价格不会上涨;特别地,当 T 的所有行和小于 1 时,则商品 1 的价格上涨,而商品 2 的价格下跌.

证明　令 $A=I-T$. 则定理条件保证 A 为非奇异 M-矩阵. 这时价格方程为

$$p_1=\sum_{i\neq 1}t_{i1}p_i+v_1+\tau,$$

$$p_2=\sum_{i\neq 2}t_{i2}p_i+v_2-\tau,$$

$$p_j=\sum_{i\neq j}t_{ij}p_i+v_j,\qquad j=3,\cdots,n,$$

亦即

$$\boldsymbol{p}=T^{\mathrm{T}}\boldsymbol{p}+(v_1+\tau,v_2-\tau,v_3,\cdots,v_n)^{\mathrm{T}}.\tag{17}$$

上式对 τ 求导数得出

$$\frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}\tau}=T^{\mathrm{T}}\frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}\tau}+(1,-1,0,\cdots,0)^{\mathrm{T}},$$

因而

$$\begin{aligned}\frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}\tau}&=(A^{\mathrm{T}})^{-1}(1,-1,0,\cdots,0)^{\mathrm{T}}\\&=(A^{-1})^{\mathrm{T}}(1,-1,0,\cdots,0)^{\mathrm{T}}.\end{aligned}$$

特别地我们应用推论 4(1)得出

$$\frac{\mathrm{d}p_1}{\mathrm{d}\tau} = (A^{-1})_{11} - (A^{-1})_{21} \geqslant 0,$$

$$\frac{\mathrm{d}p_2}{\mathrm{d}\tau} = (A^{-1})_{12} - (A^{-1})_{22} \leqslant 0.$$

此表明商品 1 的价格不会下跌,商品 2 的价格不会上涨. 当 T 的行和均小于 1 时,引理 3 保证前式右边两个严格不等式成立. 此表明:商品 1 与商品 2 的价格分别上涨与下跌. □

最后,我们通过具有三个产业部门——农业、制造业与服务业的简单经济体中产品流通情况介绍投入-产出表与技术投入-产出表,并验证本小节中有关结果.

为了从事生产活动,每个部门或产业应获得一些投入,这可能包括原材料、半成品或由别的产业买进的资产设备,同时,必须偿付商业税与雇佣劳力. 通常,某些中间产品从其他产业买进,这些产品相对于最后的或不进入这个生产过程的外部商品而言,被当作生产另一些产品与供给的投入. 由每个产业获得的产出,或卖给外面用户,或卖给其他产业部门,它们把这些产品用做投入. 在一种经济体中,综合所有不同投入的来源以及所有产业的各种各样产出销售情况的表格称为**投入-产出表**.

作为一个例子,考虑由(1)农业,(2)制造业,(3)服务业组成的假想经济体. 这些部门中每一个只生产一种产出:农产品、制造业产品或者服务供给. 它们彼此依赖,彼此从对方购进投入,又出卖产出给对方. 这里不牵连政府与对外贸易. 所有最终的产品与服务供给(它们不进入生产过程)被由消费者等组成的外部部门利用. 遵照早先提出的假定,可以给出一种表格,概括产品与服务供给的流通情况,具体如表 8.1 所示(本表选自文献[1]).

表 8.1 投入-产出表(单位:美元)

产出(到) / 投入(由)	(1) 农业	(2) 制造业	(3) 服务业	外部需求	总产出
(1)农业	15(x_{11})	20(x_{12})	30(x_{13})	35(d_1)	100(x_1)
(2)制造业	30(x_{21})	10(x_{22})	45(x_{23})	115(d_2)	200(x_2)
(3)服务业	20(x_{31})	60(x_{32})	0(x_{33})	70(d_3)	150(x_3)

这里,x_{ij} 表示部门 i 卖给部门 j 的产品量,因而

$$x_{ij} = t_{ij}x_j, \qquad 1 \leqslant i,j \leqslant 3. \tag{18}$$

上表中每行的数据表示对不同部门与用户的投入分配,而每列的数据表示为生产所需要的投入来源. 例如,从第一行数据看出,在总产出为 100 美元的农产品中,15 美元用于农业再生产,20 美元的产品卖给制造业,30 美元的产品卖给服务业,且最后 35 美元的农产品用以满足外部需求. 类似地从第二列数据看出,制造业为生产

价值 200 美元的总产出 x_2，必须投入 20 美元的农产品，10 美元的自己产品以及 60 美元的服务供给. 此时，表 8.1 的基本行关系为

$$x_i = x_{i1} + x_{i2} + x_{i3} + d_i, \qquad i = 1,2,3. \tag{19}$$

为了更便于分析，在投入-产出表基础上再造出表明每个部门为生产一个单位产出的投入需求的表格，即所谓**技术投入-产出表**，它的数据就是该经济体的投入系数 t_{ij}. 这样，由表 8.1 推算出的技术投入-产出表如表 8.2 所示：

表 8.2　技术投入-产出表

产出(到) / 投入(由)	(1) 农　业	(2) 制 造 业	(3) 服 务 业
(1)农业	0.15(t_{11})	0.10(t_{12})	0.20(t_{13})
(2)制造业	0.30(t_{21})	0.05(t_{22})	0.30(t_{23})
(3)服务业	0.20(t_{31})	0.30(t_{32})	0.00(t_{33})

表 8.2 实际上表示该模型的投入矩阵 T 为

$$T = \begin{bmatrix} 0.15 & 0.10 & 0.20 \\ 0.30 & 0.05 & 0.30 \\ 0.20 & 0.30 & 0.00 \end{bmatrix}, \tag{20}$$

于是，矩阵 $A=I-T$ 为

$$A = \begin{bmatrix} 0.85 & -0.10 & -0.20 \\ -0.30 & 0.95 & -0.30 \\ -0.20 & -0.30 & 1.00 \end{bmatrix}. \tag{21}$$

正如我们所见，表 8.1 或表 8.2 综合了这个假想经济体三个部门之间的产品买卖情况，同时，也描述了生产的技术. 例如，从表 8.1 看到，服务业 150 美元的总产出需要 30 美元的农业品与 45 美元的制造业产品的投入，它可为外部部门提供 70 美元的服务供给.

这个假想的开式模型为可行的，因为 $A \in \mathbb{Z}^{3\times3}$ 为严格对角占优矩阵，且所有 $a_{ii}>0$. 这也可以从 A 的逆正性看出，因为

$$A^{-1} = \begin{bmatrix} 1.3459 & 0.2504 & 0.3443 \\ 0.5634 & 1.2676 & 0.4930 \\ 0.4382 & 0.4304 & 1.2167 \end{bmatrix} > O, \tag{22}$$

其中所有元素舍入到四位小数. 假如 $\boldsymbol{d}=(d_1,d_2,d_3)^{\mathrm{T}}=(100,200,300)^{\mathrm{T}}$，应用 (22)式可得

$$\boldsymbol{x} = A^{-1}\boldsymbol{d} = (287.96, 457.76, 494.91)^{\mathrm{T}}. \tag{23}$$

这表明为了满足三个部门与外部部门的投入需求，需要 287.96 美元的农产品，457.76 美元的制造业产品与 494.91 美元的服务供给.

在考虑效应问题时,注意到(21)式中 A 为不可约非奇异 M-矩阵,且严格对角占优,因而引理 3 与推论 4 的假定条件均成立,这时我们有不等式(10). 现假定农业品需求单独增加到 300 美元,即有 $\hat{\boldsymbol{d}}=(300,200,300)^{\mathrm{T}}$,则新的产出向量 $\hat{\boldsymbol{x}}$ 为

$$\hat{\boldsymbol{x}} = A^{-1}\hat{\boldsymbol{d}} = (557.14, 570.44, 582.55)^{\mathrm{T}}. \tag{24}$$

比较(23)与(24)二式,我们看出,这时每个部门的产出都增加,不过农产品的产出增加量较比其他的大些,因为

$$\hat{\boldsymbol{x}} - \boldsymbol{x} = (269.18, 112.68, 87.64)^{\mathrm{T}}.$$

此与推论 6 的结果完全一致. 对于相应的价格方程组,我们也有类似于推论 9 的结果,但这里不另赘述.

习题 8.3.1

1. 试证:具有投入矩阵

$$T = \begin{bmatrix} 0.2 & 0.3 & 0.2 \\ 0.4 & 0.1 & 0.2 \\ 0.1 & 0.3 & 0.2 \end{bmatrix}$$

的开式 Leontief 模型为可行的;计算与开式需求 $d_1=10, d_2=5, d_3=6$ 相对应的总产出 x_1, x_2, x_3; 验证一下推论 6 与推论 9 可用于本题的开式模型.

2. 证明:若具有投入矩阵 T 的某开式 Leontief 模型是可行的,则 T 至少有一列元素之和小于 1.

3. 考虑一个开式 Leontief 模型,它的每个部门确切地只供应其他一个部门,且每个部门恰有一个供应者. 举例证明这样模型的投入矩阵不必为不可约的.

4. 设 $A\in Z^{n\times n}$ 为非奇异 M-矩阵. 对向量 $\boldsymbol{d}^{(1)}\geqslant\boldsymbol{0}$ 与 $\boldsymbol{d}^{(2)}\geqslant\boldsymbol{0}$,考虑方程组 $A\boldsymbol{x}^{(1)}=\boldsymbol{d}^{(1)}$ 与 $A\boldsymbol{x}^{(2)}=\boldsymbol{d}^{(2)}$. 证明:若 $\Delta\boldsymbol{d}=\boldsymbol{d}^{(2)}-\boldsymbol{d}^{(1)}$ 与 $\Delta\boldsymbol{x}=\boldsymbol{x}^{(2)}-\boldsymbol{x}^{(1)}$,则 $\boldsymbol{x}^{(1)}\geqslant\boldsymbol{0}$ 与 $\boldsymbol{x}^{(2)}\geqslant\boldsymbol{0}$,且下列断言成立:

(1) $\Delta\boldsymbol{d}\neq\boldsymbol{0}\Leftrightarrow\Delta\boldsymbol{x}\neq\boldsymbol{0}$.

(2) $\Delta\boldsymbol{d}\geqslant\boldsymbol{0}(\neq\boldsymbol{0})\Rightarrow\Delta\boldsymbol{x}\geqslant\boldsymbol{0}(\neq\boldsymbol{0})$.

(3) $\Delta\boldsymbol{d}\leqslant\boldsymbol{0}(\neq\boldsymbol{0})\Rightarrow\Delta\boldsymbol{x}\leqslant\boldsymbol{0}(\neq\boldsymbol{0})$.

并且,若 A 还是不可约的,则在(2)与(3)中分别有 $\Delta\boldsymbol{x}>\boldsymbol{0}$ 与 $\Delta\boldsymbol{x}<\boldsymbol{0}$.

考虑到开式 Leontief 模型,则上述(1)意味着当且仅当至少有一个产品外部需求变化时,则经济体中至少有一个产品的总产出变化. (2)意味着若至少有一个产品的需求增加,则至少有一产品的总产出增加. (若 A 还是不可约的,则所有产品的生产增加)(3)意味着(2)的相反情形.

5. 设 $A\in Z^{n\times n}$ 为不可约非奇异 M-矩阵. 用前题的记号,且令 $\Delta\boldsymbol{x}=(\Delta x_1,\cdots,\Delta x_n)^{\mathrm{T}}$ 与 $\Delta\boldsymbol{d}=(\Delta d_1,\cdots,\Delta d_n)^{\mathrm{T}}$. 试证:

(1) $\Delta\boldsymbol{x}>\boldsymbol{0}\Leftrightarrow\Delta x_j>0$ 对使 $\Delta d_j<0$ 的每个 j 成立.

(2) $\Delta\boldsymbol{x}<\boldsymbol{0}\Leftrightarrow\Delta x_j<0$ 对使 $\Delta d_j>0$ 的每个 j 成立.

用到开式 Leontief 模型,则(1)意味着如果所有需求减少的产品的生产增加,则所有的生产增加,而(2)意味着如果所有需求增加的产品生产减少,则所有产品的生产减少. 由于题 4 可以看出,倘若需求减少的产业的生产增加,那么该产业必定是需求增加的产业.

6. 对给定实数 α,矩阵 $A=(a_{ij})\in \boldsymbol{M}_n(\mathbb{R})$ 称为 $\mathfrak{M}(\alpha)$ 类的矩阵,假如 $a_{jj}-a_{ij}=\alpha, \forall i\neq j$. 试证:若对某实数 $\alpha, A=(a_{ij})\in\mathfrak{M}(\alpha)$,则

(1) $\det A=\alpha^{n-1}(\sum_{i=1}^{n}a_{ii}-(n-1)\alpha)$.

(2) $\alpha^{n-1}A\boldsymbol{e}=(\det A)\boldsymbol{e}$,这里 $\boldsymbol{e}=(1,\cdots,1)^{\mathrm{T}}$.

另外,当 $\alpha\neq 0$ 时,

$$A\in\mathfrak{M}(\alpha)\Leftrightarrow \mathrm{adj}A\in\mathfrak{M}((\det A)/\alpha).$$

7. 设 $A=(a_{ij})\in Z^{n\times n}$ 为满足 $A\boldsymbol{e}>\boldsymbol{0}$ 的不可约非奇异 M-矩阵. 沿用前三题的记号,试证如下命题彼此等价:

(1) 对某个实数 $\alpha, A\in\mathfrak{M}(\alpha)$.

(2) 对每个满足 $\Delta d_i\neq 0, \Delta d_j\neq 0$ 的 $\Delta\boldsymbol{d}$ 以及每对 i 与 j, $\Delta x_i<\Delta x_j\Leftrightarrow\Delta d_i<\Delta d_j$.

(3) 对每个 $\Delta\boldsymbol{d}\neq\boldsymbol{0}, \Delta x_m=\max_i\Delta x_i\Leftrightarrow\Delta d_m=\max_i\Delta d_i$.

(4) 对每个 $\Delta\boldsymbol{d}\neq\boldsymbol{0}, \Delta x_m=\min_i\Delta x_i\Leftrightarrow\Delta d_m=\min_i\Delta d_i$.

按开式 Leontief 模型的语言,(3)与(4)断定若投入矩阵 T 不可约且所有行和皆小于 1,则需求与生产的最大或最小变化总出现在同一产业中. 当且仅当对某实数 $\alpha, A\in\mathfrak{M}(\alpha)$ 时,这里 $A=I-T$,才会发生(3)与(4)的现象. 还要注意,本题与定理 5、定理 8 密切相关.

8. 对投入矩阵 T

$$T=\begin{bmatrix}0 & \frac{3}{10} & \frac{3}{5}\\ \frac{1}{5} & \frac{1}{10} & \frac{3}{5}\\ \frac{1}{5} & \frac{3}{10} & \frac{2}{5}\end{bmatrix}$$

证明:存在实数 α 使得 $A=I-T\in\mathfrak{M}(\alpha)$. 并对 $\Delta\boldsymbol{d}=(12,-12,24)^{\mathrm{T}}$ 计算 $\Delta\boldsymbol{x}$,以此来验证前题(2),(3)与(4)的结论此时对这一对 $\Delta\boldsymbol{x},\Delta\boldsymbol{d}$ 成立.

9. 设 $T\in\mathcal{N}_n$ 除了第 k 行外,其余各行和严格小于 s. 令 $C=(c_{ji})=\mathrm{adj}(sI-T)$. 试证:

$$c_{kk}>c_{kj},\qquad \forall j\neq k.$$

10. 设 $C=(c_{ji})=\mathrm{adj}(sI-T), s>0, T\in\mathcal{N}_n, s>\rho(T)$. 试证:对 $n\geqslant 3, c_{ii}c_{kj}-c_{ij}c_{ki}\geqslant 0, i,j,k=1,\cdots,n$.

11. 证明:若 $A\in\boldsymbol{M}_n(\mathbb{R})$(行)严格对角占优,则 A^{-1} 的元素 $(A^{-1})_{ij}$ 满足条件: $|(A^{-1})_{ii}|>|(A^{-1})_{ji}|, \forall j\neq i$.

8.3.2　闭式 Leontief 模型

在 8.3.1 小节中,我们考虑开式投入-产出模型,这时,开式部门在经济体之外. 假若将这个外部部门吸收到经济体里作为向其他部门提供劳力的一个部门,而其所得到其他部门的产品则作为维持生存的工资组成部分,则在这样的模型中,最终需求与原先的投入不再出现,代替它们的将是刚才提到的新产业部门投入需求与产出. 由于生产出的任何东西只是为满足模型中部门或产业本身的投入需求,

所有产品皆为中间产品. 这样的投入-产出模型称为**闭式 Leontief 模型**或**闭式模型**. 在这种系统里,原先的最终需求以及诸如雇佣与工资率等项目都当做未知量来处理,且它们的平衡值将同其他变量一块地被解出.

由于没有变量由外部确定,闭式模型的基本问题变为:在给定生产技术条件下,怎么样的平衡产出与价格水平才能满足各种需求? 为分析这一问题,我们先从一个看来与 8.3.1 小节方程组(2),即 $A\boldsymbol{x}=\boldsymbol{d}(A=I-T)$类似的方程组入手,这里,$x_i$ 为部门 i 的产出,d_i 现在与雇佣水平 ε 有关,即消费者购买一个量 d_i 是随着 ε 而变化的. 我们用 $c_i=d_i/\varepsilon$ 表示固定的技术系数,因而 $d_i=c_i\varepsilon$. 假定原先开式模型含有 $n-1$ 个部门,则在闭式模型里,投入-产出平衡关系由 $n-1$ 个联立线性方程

$$x_i=\sum_{j=1}^{n-1}t_{ij}x_j+c_i\varepsilon,\qquad i=1,\cdots,n-1 \tag{1}$$

来描述. 现在全部劳力的供应,即雇佣水平,为每个部门使用劳力之和:

$$\varepsilon=\sum_{i=1}^{n}L_i,$$

其中,L_i 代表部门 i 使用的劳力数,且这里 $i=n$ 为消费者部门. 今用

$$l_i=L_i/x_i,\quad 1\leqslant i\leqslant n-1;\quad l_n=L_n/\varepsilon$$

定义固定的劳力投入系数. 那么,全部劳力供应可用下列线性方程表示

$$\varepsilon=\sum_{j=1}^{n-1}l_jx_j+l_n\varepsilon. \tag{2}$$

联立(1)与(2)两式,我们有

$$\begin{aligned}
x_1&=t_{11}x_1+t_{12}x_2+\cdots+t_{1n-1}x_{n-1}+c_1\varepsilon,\\
x_2&=t_{21}x_1+t_{22}x_2+\cdots+t_{2n-1}x_{n-1}+c_2\varepsilon,\\
&\ \vdots\\
x_{n-1}&=t_{n-11}x_1+t_{n-12}x_2+\cdots+t_{n-1n-1}x_{n-1}+c_{n-1}\varepsilon,\\
\varepsilon&=l_1x_1+l_2x_2+\cdots+l_{n-1}x_{n-1}+l_n\varepsilon,
\end{aligned} \tag{3}$$

为了便于书写,我们令 $T=(t_{ij})\in\boldsymbol{M}_n(\mathbb{R})$,其中

$$t_{in}=c_i,\quad 1\leqslant i\leqslant n-1;\quad t_{nj}=l_j,\quad 1\leqslant j\leqslant n.$$

$$\boldsymbol{x}=(x_1,x_2,\cdots,x_{n-1},x_n)^{\mathrm{T}},\qquad x_n=\varepsilon.$$

则方程组(3)可以写为向量-矩阵形式

$$\boldsymbol{x}=T\boldsymbol{x}. \tag{4}$$

这里,$T=(t_{ij})$叫作该闭式模型的**投入矩阵**. 当令 $A=I-T$ 时,(4)式可改写为

$$A\boldsymbol{x}=\boldsymbol{0}. \tag{5}$$

由于(5)式为齐次线性方程组,故它或只有平凡解 $\boldsymbol{x}=\boldsymbol{0}$,或有无穷多个解. 如前,我们的兴趣在于它的非负解 $\boldsymbol{x}$.

从前面看出,闭式模型顾及到需求对供应影响以及供应对需求影响两个方面,

这由同时求解 ε 与产出供应 $x_j(1\leqslant i\leqslant n-1)$ 的办法来实现.

仍用 $\boldsymbol{x}$ 表示闭式模型的产出向量,则由于 $\sum_{j=1}^{n} t_{ij}x_j$ 为生产这些产出所需要的第 i 个投入量,故 $\boldsymbol{y}=T\boldsymbol{x}$ 表示投入向量. 但因诸产出 x_i 是系统投入的仅有的来源,所以除了 $\boldsymbol{y}\leqslant\boldsymbol{x}$ 以外,整个系统不可能运转. 而且,生产过程被假定为不可逆转的,于是 $\boldsymbol{x}$ 一定为非负的,且实际上要求它为正向量. 这引出如下定义.

定义 1 具有投入矩阵 T 的某闭式 Leontief 模型称为**可行的**,假如存在 $\boldsymbol{x}>\boldsymbol{0}$,使得 $T\boldsymbol{x}\leqslant\boldsymbol{x}$. 若这样的 $\boldsymbol{x}$ 存在,则称之为闭式模型的**可行产出解**.

当某闭式模型可行时,我们希望找出满足 $T\boldsymbol{x}=\boldsymbol{x}$ 的可行产出解 $\boldsymbol{x}>\boldsymbol{0}$. 这样的 $\boldsymbol{x}$ 称为该闭式模型的一个**均衡产出向量**. 因此,某闭式模型(以 T 为投入矩阵)有一个均衡产出向量当且仅当齐次方程组(5)有正向量解 $\boldsymbol{x}>\boldsymbol{0}$. 但要注意一个可行的闭式模型不一定拥有一个均衡产出向量. 例如,对应于下列投入矩阵的闭式模型虽然可行,但没有均衡产出向量:

$$T=\begin{bmatrix}0&0\\0&1\end{bmatrix}. \tag{6}$$

现在考虑闭式模型可行性的有关条件以及均衡产出向量存在性问题. 在这里具有"性质 c"的 M-矩阵理论起着基本作用.

定理 2 若具有投入矩阵 T 的某闭式 Leontief 模型是可行的,则 $A=I-T$ 为具有"性质 c"的 M-矩阵.

证明 这是 8.2.2 小节定理 4 的直接推论. □

应用 8.2.2 小节定理 3,我们有如下结论.

推论 3 若具有投入矩阵 T 的某闭式 Leontief 模型是可行的,令 $A=I-T$,则存在一个对称正定矩阵 W 使得 $AW+WA^{\mathrm{T}}$ 为对称正半定的.

我们指出定理 2 中条件只是必要的不是充分的. 例如,

$$A=I-T=\begin{bmatrix}1&0\\0&1\end{bmatrix}-\begin{bmatrix}1&1\\0&0\end{bmatrix}=\begin{bmatrix}0&-1\\0&1\end{bmatrix}$$

为具有"性质 c"的 M-矩阵,但不存在 $\boldsymbol{x}=(x_1,x_2)^{\mathrm{T}}>\boldsymbol{0}$ 使得 $A\boldsymbol{x}\geqslant\boldsymbol{0}$.

为了进一步研究 Leontief 模型的可行性与均衡产出向量的存在性问题,我们考虑 $T\in\mathcal{N}_n$ 的可约法式:

$$PTP^{\mathrm{T}}=\begin{bmatrix}T_{11}&T_{12}&\cdots&T_{1k}\\&T_{22}&\cdots&T_{2k}\\&0&\ddots&\vdots\\&&&T_{kk}\end{bmatrix}, \tag{7}$$

这里,每个 T_{ii} 或为不可约矩阵或为一阶零矩阵. 为了我们这里的方便,也可将一阶零矩阵视为不可约矩阵. 同时,当 T 不可约时,简单地记 $T=[T_{11}]$.

引理 4　设某闭式模型的投入矩阵 T 为不可约的，且 $A=I-T$ 奇异. 则下列诸命题彼此等价：

(1) 该模型是可行的，即 $A\boldsymbol{x}\geqslant\mathbf{0}$ 与 $\boldsymbol{x}>\mathbf{0}$ 相容.

(2) 该模型有一个均衡产出向量 $\boldsymbol{x}$，如不计正数因子，它是唯一的.

(3) A 为具有“性质 c”的 M-矩阵.

证明　(1)⇒(3)：见定理 2. (3)⇒(2)：设(3)成立，则 $A=I-T$ 为奇异不可约 M-矩阵，因而根据 8.2.1 小节定理 11(2)，存在 $\boldsymbol{x}>\mathbf{0}$ 满足 $A\boldsymbol{x}=\mathbf{0}$. 但因 T 不可约，所以有 $\rho(T)=1$ 与 $T\boldsymbol{x}=\boldsymbol{x}$，按 Perron-Frobenius 定理，如不计正数因子，$\boldsymbol{x}$ 为唯一的. 因此，(2)成立. 最后，(2)⇒(1)直接由定义推出. □

当投入矩阵 T 为可约矩阵且有可约法式(7)时，以下重要结论指出，对应的闭式 Leontief 模型可行当且仅当 $A=I-T$ 为 M-矩阵，且若 PAP^{T} 与(7)式中 PTP^{T} 有一致的分块矩阵形式，则对每一个 $i,1\leqslant i\leqslant k,A_{ii}$ 奇异蕴涵 $T_{ij}=O,\forall j\neq i$(见第七章习题 7.2.2 题 5).

定理 5　设 $T\in\mathcal{N}_n$ 为某闭式 Leontief 模型的投入矩阵，其可约法式如(7)式. 令 $A=I-T$，且 PAP^{T} 与 PTP^{T} 有一致的分块矩阵形式. 则该模型是可行的，当且仅当下列两条件同时成立：

(1) A 为 M-矩阵.

(2) 对每一个 $i,1\leqslant i\leqslant k$，

$$A_{ii}\text{ 奇异}\Rightarrow T_{ij}=O,\forall j\neq i. \tag{8}$$

证明　设模型是可行的. 则由定理 2 与定义 1 知道，$A=I-T$ 为 M-矩阵，且存在一个分块向量 $\boldsymbol{x}=(\boldsymbol{y}_1^{\mathrm{T}},\cdots,\boldsymbol{y}_k^{\mathrm{T}})^{\mathrm{T}}>\mathbf{0}$ 使得 $PTP^{\mathrm{T}}\boldsymbol{x}\leqslant\boldsymbol{x}$. 由于 $A_{ij}=\delta_{ij}I-T_{ij},i,j=1,\cdots,k$，我们有

$$\sum_{j\geqslant i}A_{ij}\boldsymbol{y}_j\geqslant\mathbf{0},\qquad\forall\,1\leqslant i\leqslant k. \tag{9}$$

但 $A_{ij}\leqslant O,\forall j>i$，上式推出，对所有 $1\leqslant i\leqslant k,A_{ii}\boldsymbol{y}_i\geqslant\mathbf{0}$. 因 A_{ii} 不可约，所以若 A_{ii} 奇异(即 A_{ii} 为奇异不可约 M-矩阵)，则根据 8.2.1 小节定理 11(4)，$A_{ii}\boldsymbol{y}_i\geqslant\mathbf{0}$ 蕴涵 $A_{ii}\boldsymbol{y}_i=\mathbf{0}$，因而(9)式推出 $A_{ij}=O,\forall j\neq i$，此等价于 $T_{ij}=-A_{ij}=O,\forall j\neq i$.

反之，设(1)与(2)成立. 当 A_{ii} 奇异时，它必为不可约奇异 M-矩阵. 根据 8.2.1 小节定理 11(2)，存在 $\boldsymbol{y}_i>\mathbf{0}$ 使得 $A_{ii}\boldsymbol{y}_i=\mathbf{0}$. 另一方面，当 A_{ii} 非奇异时，则 A_{ii} 为非奇异 M-矩阵，因而按 8.1.2 小节定理 3(2)，存在 $\boldsymbol{y}_i>\mathbf{0}$ 使得 $A_{ii}\boldsymbol{y}_i>\mathbf{0}$. 因此，由于条件(2)，又由于可以适当地缩小诸 $\boldsymbol{y}_i$ 使得 $\boldsymbol{x}=(\boldsymbol{y}_1^{\mathrm{T}},\cdots,\boldsymbol{y}_k^{\mathrm{T}})^{\mathrm{T}}>\mathbf{0}$，且(9)式成立，我们有 $PAP^{\mathrm{T}}\boldsymbol{x}\geqslant\mathbf{0}$ 亦即 $PTP^{\mathrm{T}}\boldsymbol{x}\leqslant\boldsymbol{x}$. 由此推得，$T(P^{\mathrm{T}}\boldsymbol{x})\leqslant P^{\mathrm{T}}\boldsymbol{x},P^{\mathrm{T}}\boldsymbol{x}>\mathbf{0}$，按定义 1，该模型为可行的. □

正如前面指出的，可行的闭式 Leontief 模型不一定有一个均衡产出向量，但如在前一定理条件(2)中再加上一个条件，便可得到有均衡产出向量的可行闭式模型

的一个重要特征. 确切地说我们有

定理 6　设 A 与 T 同定理 5. 则对应投入矩阵 T 的闭式 Leontief 模型有一个均衡产出向量当且仅当下列两条件同时成立：

(1) A 为 M-矩阵.

(2) 对每一个 i, $1 \leqslant i \leqslant k$,

$$A_{ii} \text{ 奇异} \Leftrightarrow T_{ij} = O, \qquad \forall j \neq i. \tag{10}$$

证明　这是第七章 7.2.2 小节引理 4 的直接推论. □

现在讨论闭式模型的价格方程. 由于不存在剩余，价格就等于生产成本，即有

$$p_j = \sum_{i=1}^{n} p_i t_{ij}, \qquad j = 1, \cdots, n.$$

用 $\boldsymbol{p} = (p_1, \cdots, p_n)^{\mathrm{T}}$ 表示价格向量，则上式可写成

$$\boldsymbol{p} = T^{\mathrm{T}} \boldsymbol{p}, \tag{11}$$

或等价地

$$A^{\mathrm{T}} \boldsymbol{p} = \boldsymbol{0}. \tag{12}$$

类似于产出方程情形，我们将方程组(12)的正向量解 $\boldsymbol{p}$ 称为对应闭式 Leontief 模型的**均衡价格向量**. 均衡价格向量存在性是经济均衡理论的基本问题之一. 在这里的闭式模型中，应用定理 6 可推得该模型有均衡价格向量的一个充分与必要条件.

推论 7　设 A 与 T 同定理 5. 则对应投入矩阵 T 的闭式 Leontief 模型有一个均衡价格向量当且仅当下列两条件同时成立：

(1) A 为 M-矩阵.

(2) 对每一个 i, $1 \leqslant i \leqslant k$,

$$A_{ii} \text{ 奇异} \Leftrightarrow T_{ji} = O, \qquad \forall j \neq i. \tag{13}$$

最后我们指出，如果一个可行闭式 Leontief 模型的投入矩阵 T 为不可约的，那么均衡产出向量 $\boldsymbol{x}$ 也可用有限 Markov 链理论中计算相应于一个遍历 Markov 链的平稳分布向量的方法得出.

习题 8.3.2

1. 考虑投入-产出关系由下列方程组给出的闭式 Leontief 模型：

$$x_1 = 0.2x_1 + 0.2x_2,$$
$$x_2 = 0.3x_2 + 0.3\varepsilon,$$
$$\varepsilon = 0.8x_1 + 0.5x_2 + 0.7\varepsilon.$$

这里，x_1 与 x_2 为两个内部部门的产出，ε 为雇佣水平. 试证上述模型为可行的，且它有均衡产出向量 $(x_1, x_2, \varepsilon)^{\mathrm{T}}$，如不计正数因子，此向量是唯一的. 试在 $x_1 + x_2 + \varepsilon = 1$ 约束下算出此向量.

2. 试给出一个闭式 Leontief 模型的例子，它至少有两个线性相关的均衡产出向量.

3. 试证：具有对称投入矩阵 T 的闭式 Leontief 模型是可行的，当且仅当 $A = I - T$ 为 M-矩阵. 但是，这样的模型不必有一个均衡产出向量. 试用 T 的可约法式(7)给出均衡产出向量存在

的充分与必要条件.

参考文献

[1] Berman A, Plemmons R J. Nonnegative Matrices in the Mathematical Sciences. New York: Academic Press, 1979, Reprinted and updated, Philadelphia: SIAM Press, 1994

[2] Seneta E. Non-negative Matrices. New York: Wiley, 1973

[3] Fiedler M, Pták V. On matrices with nonpositive off-diagonal elements and positive principal minors. Czech. Math. J., 1962(12), 382～400

[4] Hoffman A J. Generalizations of Geršgorin's theorem: G-generating familices. Lecture notes, Univ. of California at Santa Barbara. Aug. 1969

[5] Ky Fan. Note on circular disks containing the eigenvalues of a matrix. Duke Math. J., 1958(25), 411～445

[6] Beauwens R. Semistrict diagonal dominance. SIAM J. Numer. Anal., 1976(13), 104～112

[7] Berman A, Varga R S, Ward R C. ALPS: Matrices with nonpositive off-diagonal entries. Linear Algebra Appl., 1978(21), 233～244

[8] Carlson D H, Varga R S. Minimal G-functions. Linear Algebra Appl., 1973(6), 97～117

[9] Chen G N. (陈公宁). Some generalizations of diagonal dominance associated with G-functions. Linear Algebra Appl., 1983(55), 169～180

[10] Chen G N. (陈公宁). Note on lower bounds for the rank of a matrix. Linear Algebra Appl., 1983(55): 125～132

[11] Meyer C D, Plemmons R J. Convergent powers of a matrix with applications to iterative methods for singular linear systems. SIAM J. Numer. Anal., 1977(14): 699～705

[12] Neumann M, Plemmons R J. Convergent nonnegative matrices and iterative methods for consistent linear systems. Numer. Math., 1978(31), 265～279

[13] Neumann M, Plemmons R J. Generalized inverse-positivity and splittings of M-matrices. Linear Algebra Appl, 1979(23), 21～26

[14] Plemmons R J. M-matrices leading to semiconvergent splittings. Linear Algebra Appl., 1976(15), 243～252

[15] Rothblum U G. An index classification of M-matrices. Linear Algebra Appl., 1979(23), 1～12

[16] Varga R S. On recurring theorems on diagonal dominance. Linear Algebra Appl., 1976(13), 1～9

[17] Stern R J, Tsatsomeros M. Extended M-matrices and subtangentiality. Linear Algebra Appl., 1987(97), 1～11

[18] Kuo I-wen. A note on factorizations of singular M-matrices. Linear Algebra Appl., 1977(16), 217～220

[19] Meyer C D. The role of the group generalized inverse in the theory of finite Markov chains. SIAM Rev., 1975(17), 443～464

[20] Sierksma G. Non-negative matrices; the open Leontief model . Linear Algebra Appl. , 1979(26),175～201
[21] Varga R S. Matrix Iterative Analysis. 2nd ed. ,Berlin:Springer. 2000
[22] Yan C. Introduction to Input -Output Economics. New York:Holt,1969
[23] Leontief W W. Quantitative input and output relations in the economic system of the United States. Rev. Econ. Statist. ,1936(18),100～125
[24] Leontief W W. Input-Output Economics. London and New York:Oxford Univ. Press,1966
[25] Sierksma G, Bakker E J. Nearly principal minors of M-matrices. Compositio Math . , 1986(59),73～79
[26] Takayama A. Mathematical Econmics. Dryden,Hinsdale,Illinois,1974
[27] 华罗庚. 计划经济大范围最优化的数学理论,(I)－(IX). 科学通报,1984(12),705～709; 1984(13),769～772;1984(16),961～965;1984(18),1089～1092;1984(21),1281～1282; 1985(1),1～2
[28] 陈公宁. 弱半正矩阵与弱对角稳定矩阵的一些特性. 数学杂志,1988(8),263～268

符　号　表

符号	含义
$\{1,2,3\}$	元素 1,2,3 的有限集合
$\mathbb{R},\mathbb{R}_+$	实数集合,非负实数集合
$\mathbb{R}^n,\boldsymbol{M}_{n,1}(\mathbb{R})$	实 n 维列向量的实向量空间
$\mathbb{C}$	复数集合
$\mathbb{C}^n,\boldsymbol{M}_{n,1}(\mathbb{C})$	复 n 维列向量的复向量空间
$\boldsymbol{F}$	域(通常为$\mathbb{R}$或$\mathbb{C}$)
$\boldsymbol{F}^n,\boldsymbol{M}_{n,1}(\boldsymbol{F})$	元素取在域 $\boldsymbol{F}$ 的 n 维列向量组成的域 F 上的向量空间
$\boldsymbol{M}_{m,n}(\boldsymbol{F})$	元素取在 $\boldsymbol{F}$ 的 $m\times n$ 矩阵的集合
$\boldsymbol{M}_{m,n}(\mathbb{R})$	$m\times n$ 实矩阵的集合
$\boldsymbol{M}_n(\mathbb{R})$	n 阶实矩阵的集合
$\boldsymbol{M}_{m,n}(\mathbb{C})$	$m\times n$ 复矩阵的集合
$\boldsymbol{M}_n(\mathbb{C})$	n 阶复矩阵的集合
A,B,C,等	矩阵;$A=(a_{ij})\in\boldsymbol{M}_{m,n}(\boldsymbol{F})$
$\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}$,等	列向量;$\boldsymbol{x}=(x_1,\cdots,x_n)^{\mathrm{T}}\in\boldsymbol{F}^n$
$\boldsymbol{0}$	零向量
$O,O_{m,n},O_n$	零矩阵,$m\times n$ 零矩阵,n 阶零矩阵
I,I_n	单位矩阵,n 阶单位矩阵
A^{T}	$A\in\boldsymbol{M}_{m,n}(\boldsymbol{F})$的转置矩阵
A^*	$A\in\boldsymbol{M}_{m,n}(\mathbb{C})$的共轭转置矩阵
A^{-1}	非奇异矩阵 $A\in\boldsymbol{M}_n(\boldsymbol{F})$的逆
$A^{1/2}$	Hermite 正半定矩阵 A 的唯一 Hermite 正半定平方根
$\|A\|$	$A\in\boldsymbol{M}_{m,n}(\boldsymbol{F})$元素加模后的矩阵
$A^{\dagger}$	$A\in\boldsymbol{M}_{m,n}(\mathbb{C})$的 Moore-Penrose 逆
$A^{\#}$	$A\in\boldsymbol{M}_n(\mathbb{C})$的群逆
A^{D}	$A\in\boldsymbol{M}_n(\mathbb{C})$的 Drazin 逆
$A\{1,2\}$等	$A\in\boldsymbol{M}_{m,n}(\mathbb{C})$的$\{1,2\}$-逆的集合等
$\mathrm{tr}A$	$A\in\boldsymbol{M}_n(\boldsymbol{F})$的迹
$\det A$	$A\in\boldsymbol{M}_n(\boldsymbol{F})$的行列式
$\mathrm{per}A$	$A\in\boldsymbol{M}_n(\boldsymbol{F})$的积和式
$\rho(A)$	$A\in\boldsymbol{M}_n(\mathbb{C})$的谱半径
$\sigma(A)$	$A\in\boldsymbol{M}_n(\mathbb{C})$的谱
$\mathrm{rank}A$	$A\in\boldsymbol{M}_n(\boldsymbol{F})$的秩
$\mathrm{adj}A$	$A\in\boldsymbol{M}_n(\boldsymbol{F})$的伴随矩阵
$m_A(\lambda)$或 $m(\lambda)$	$A\in\boldsymbol{M}_n(\mathbb{C})$的最小多项式
$A\circ B$	$A,B\in\boldsymbol{M}_{m,n}(\boldsymbol{F})$的 Hadamard 积
$A\otimes B$	A,B 的张量积或 Kronecker 积
$\mathrm{Im}A$	$A\in\boldsymbol{M}_{m,n}(\boldsymbol{F})$的值域或列空间
$\mathrm{Ker}A$	$A\in\boldsymbol{M}_{m,n}(\boldsymbol{F})$的核或零空间
$\{\lambda_i(A)\}$	$A\in\boldsymbol{M}_n(\mathbb{C})$的特征值列(计入重数)
$\{s_i(A)\}$	$A\in\boldsymbol{M}_{m,n}(\boldsymbol{F})$的奇异值列(计入重数)
δ_{ij}	Kronecker 符号
$A[\alpha\|\beta]$	$A\in\boldsymbol{M}_{m,n}(\boldsymbol{F})$的子阵
$\mathrm{codet}A[\alpha\|\beta]$	$A\in\boldsymbol{M}_n(\boldsymbol{F})$的主子式 $\det A[\alpha\|\beta]$的余子式
$[A/A[\alpha]]$	$A\in\boldsymbol{M}_n(\boldsymbol{F})$内 $A[\alpha]$的 Schur 余量
$\boldsymbol{e}_i$	$\boldsymbol{F}^n$ 中第 i 个单位坐标向量
$[\boldsymbol{x}]_{\mathscr{E}}$	向量 $\boldsymbol{x}$ 关于基 $\mathscr{E}$ 的表示
${}_{\mathscr{G}}[T]_{\mathscr{E}}$	线性变换 T 关于基对$(\mathscr{E},\mathscr{G})$的表示
$\oplus,\dotplus$	正交和或张量和,直和
$\langle\cdot,\cdot\rangle$或$(\cdot,\cdot)$	内积
$\\|\cdot\\|,\\|\cdot\\|_D$	范数,对偶范数

符号	含义
Span$S$	向量空间子集 S 的生成集
$\mathrm{sgn}(\sigma)$	置换 σ 的符号
$\mathrm{diag}[d_1,\cdots,d_n]$	对角元素为 $d_1,\cdots,d_n$ 的分块对角矩阵
$\mathrm{diag}[D_1,\cdots,D_m]$	对角块为 $D_1,\cdots,D_m$ 的分块对角矩阵
$\mathrm{In}A,\mathrm{In}^0A,\mathrm{In}'A$	$A\in \boldsymbol{M}_n(\mathbb{C})$的惯性
$\kappa(A)$	$A\in \boldsymbol{M}_n(\boldsymbol{F})$的条件数
$\mathrm{lub}_S(A),\mathrm{glb}_S(A)$	$A\in \boldsymbol{M}_n(\mathbb{C})$对于均衡凸体 S 的界与下界
$W(A),r(A)$	$A\in \boldsymbol{M}_n(\mathbb{C})$的数值域与数值半径
$\mathscr{U}_n$	$M_n(\mathbb{C})$中酉矩阵的集合
$\dim\mathscr{S}$	线性空间 $\mathscr{S}$的维数
vec$A$	$A\in \boldsymbol{M}_{m,n}(\mathbb{C})$的向量函数
sig$A$	$A\in \boldsymbol{M}_n(\mathbb{C})$的正负号差
$\mathrm{Bez}(a,b)$	复多项式对 $a(\lambda),b(\lambda)$的 Bézout 矩阵
$R(a,b)$	复多项式对 $a(\lambda),b(\lambda)$的结式矩阵
$\mathrm{index}(A)$	$A\in \boldsymbol{M}_n(\mathbb{C})$的指标
ReA, ImA	$A\in \boldsymbol{M}_n(\mathbb{C})$的实部与虚部
$\Gamma(A)$	$A\in \boldsymbol{M}_n(\mathbb{C})$的有向图
$\gamma(A)$	素矩阵 A 的素性指标
$\mathscr{N}_n$	n 阶非负矩阵集合
Ω_n	n 阶双随机矩阵的集合
$\mathscr{L}(\mathscr{S}_1,\mathscr{S}_2)$	自向量空间 $\mathscr{S}_1$ 到 $\mathscr{S}_2$ 的线性变换的集合
$\delta(A)$	$\max\{\lvert\lambda\rvert:\lambda\in\sigma(A),\lambda\neq\rho(A)\}$
$\mathscr{M}(A)$	$A\in \boldsymbol{M}_n(\mathbb{C})$的比较矩阵
S_n	n 阶对称群
$\in$	元素属于
$\subset$	集合含于
$\cup$	集合的并
$\cap$	集合的交
$\varnothing$	空集
$\Rightarrow$	蕴涵
$\Leftrightarrow$	等价
$\square$	证明或求解结束记号

《现代数学基础丛书》出版书目

（按出版时间排序）

1 数理逻辑基础(上册) 1981.1 胡世华、陆钟万 著
2 紧黎曼曲面引论 1981.3 伍鸿熙、吕以辇、陈志华 著
3 组合论(上册) 1981.10 柯召、魏万迪 著
4 数理统计引论 1981.11 陈希孺 著
5 多元统计分析引论 1982.6 张尧庭、方开泰 著
6 概率论基础 1982.8 严士健、王隽骧、刘秀芳 著
7 数理逻辑基础(下册)1982.8 胡世华、陆钟万 著
8 群构造(上册) 1982.11 张远达 著
9 有限群构造(下册) 1982.12 张远达 著
10 环与代数 1983.3 刘绍学 著
11 测度论基础 1983.9 朱成熹 著
12 分析概率论 1984.4 胡迪鹤 著
13 巴拿赫空间引论 1984.8 定光桂 著
14 微分方程定性理论 1985.5 张芷芬、丁同仁、黄文灶、董镇喜 著
15 傅里叶积分算子理论及其应用 1985.9 仇庆久、陈恕行 编
16 辛几何引论 1986.3 J.柯歇尔、邹异明 著
17 概率论基础和随机过程 1986.6 王寿仁 著
18 算子代数 1986.6 李炳仁 著
19 线性偏微分算子引论(上册) 1986.8 齐民友 著
20 实用微分几何引论 1986.11 苏步青、华宣积、忻云龙 著
21 微分动力系统原理 1987.2 张筑生 著
22 线性代数群表示导论(上册)1987.2 曹锡华、王建磐 著
23 模型论基础 1987.8 王世强 著
24 递归论 1987.11 莫绍揆 著
25 有限群导引(上册) 1987.12 徐明曜 著
26 组合论(下册) 1987.12 柯召、魏万迪 著
27 拟共形映射及其在黎曼曲面论中的应用 1988.1 李忠 著
28 代数体函数与常微分方程 1988.2 何育赞 著
29 同调代数 1988.2 周伯壎 著
30 近代调和分析方法及其应用 1988.6 韩永生 著
31 带有时滞的动力系统的稳定性 1989.10 秦元勋等 编著

32 代数拓扑与示性类 1989.11 马德森著 吴英青、段海鲍译
33 非线性发展方程 1989.12 李大潜、陈韵梅 著
34 反应扩散方程引论 1990.2 叶其孝、李正元 著
35 仿微分算子引论 1990.2 陈恕行、仇庆久、李成章 编
36 公理集合论导引 1991.1 张锦文 著
37 解析数论基础 1991.2 潘承洞、潘承彪 著
38 拓扑群引论 1991.3 黎景辉、冯绪宁 著
39 二阶椭圆型方程与椭圆型方程组 1991.4 陈亚浙、吴兰成 著
40 黎曼曲面 1991.4 吕以辇、张学莲 著
41 线性偏微分算子引论(下册) 1992.1 齐民友 著
42 复变函数逼近论 1992.3 沈燮昌 著
43 Banach 代数 1992.11 李炳仁 著
44 随机点过程及其应用 1992.12 邓永录、梁之舜 著
45 丢番图逼近引论 1993.4 朱尧辰、王连祥 著
46 线性微分方程的非线性扰动 1994.2 徐登洲 马如云 著
47 广义哈密顿系统理论及其应用 1994.12 李继彬、赵晓华、刘正荣 著
48 线性整数规划的数学基础 1995.2 马仲蕃 著
49 单复变函数论中的几个论题 1995.8 庄圻泰 著
50 复解析动力系统 1995.10 吕以辇 著
51 组合矩阵论 1996.3 柳柏濂 著
52 Banach 空间中的非线性逼近理论 1997.5 徐士英、李冲、杨文善 著
53 有限典型群子空间轨道生成的格 1997.6 万哲先、霍元极 著
54 实分析导论 1998.2 丁传松等 著
55 对称性分岔理论基础 1998.3 唐云 著
56 Gel'fond-Baker 方法在丢番图方程中的应用 1998.10 乐茂华 著
57 半群的 S-系理论 1999.2 刘仲奎 著
58 有限群导引(下册) 1999.5 徐明曜 著
59 随机模型的密度演化方法 1999.6 史定华 著
60 非线性偏微分复方程 1999.6 闻国椿 著
61 复合算子理论 1999.8 徐宪民 著
62 离散鞅及其应用 1999.9 史及民 编著
63 调和分析及其在偏微分方程中的应用 1999.10 苗长兴 著
64 惯性流形与近似惯性流形 2000.1 戴正德、郭柏灵 著
65 数学规划导论 2000.6 徐增堃 著
66 拓扑空间中的反例 2000.6 汪林、杨富春 编著
67 拓扑空间论 2000.7 高国士 著
68 非经典数理逻辑与近似推理 2000.9 王国俊 著
69 序半群引论 2001.1 谢祥云 著

70 动力系统的定性与分支理论 2001.2 罗定军、张祥、董梅芳 编著
71 随机分析学基础（第二版）2001.3 黄志远 著
72 非线性动力系统分析引论 2001.9 盛昭瀚、马军海 著
73 高斯过程的样本轨道性质 2001.11 林正炎、陆传荣、张立新 著
74 数组合地图论 2001.11 刘彦佩 著
75 光滑映射的奇点理论 2002.1 李养成 著
76 动力系统的周期解与分支理论 2002.4 韩茂安 著
77 神经动力学模型方法和应用 2002.4 阮炯、顾凡及、蔡志杰 编著
78 同调论——代数拓扑学之一 2002.7 沈信耀 著
79 金兹堡-朗道方程 2002.8 郭柏灵、黄海洋、蒋慕蓉 著
80 排队论基础 2002.10 孙荣恒、李建平 著
81 算子代数上线性映射引论 2002.12 侯晋川、崔建莲 著
82 微分方法中的变分方法(修订版) 2003.2 陆文端 著
83 周期小波理论及其应用 2003.3 彭思龙、李登峰、谌秋辉 著
84 集值分析 2003.8 李雷 吴从炘 著
85 数理逻辑引论与归结原理 2003.8 王国俊 著
86 强偏差定理与分析方法 2003.8 刘文 著
87 椭圆与抛物型方程引论 2003.9 伍卓群、尹景学 著
88 有限典型群子空间轨道生成的格(第二版) 2003.10 万哲先、霍元极 著
89 调和分析及其在偏微分方程中的应用(第二版) 2004.3 苗长兴 著
90 稳定性和单纯性理论 2004.6 史念东 著
91 发展方程数值计算方法 2004.6 黄明游 编著
92 传染病动力学的数学建模与研究 马知恩 周义仓 王稳地 靳祯 著
93 模李超代数 2004.9 张永正、刘文德 著
94 巴拿赫空间中算子广义逆理论及其应用 2005.1 王玉文著
95 巴拿赫空间结构和算子理想 2005.3 钟不杰 著
96 脉冲微分系统引论 2005.3 傅希林、闫宝强、刘衍胜 著
97 代数学中的 Frobensius 结构 2005.7 汪明义 著
98 生存数据统计分析 2005.12 王启华 著
99 数量逻辑引论与归结原理(第二版) 2006.3 王国俊 著
100 数据包络分析 2006.3 魏权龄 著
101 代数群引论 2006.9 黎景辉 陈志杰 赵春来 著
102 矩阵结合方案 2006.9 王仰贤 霍元极 麻常利 著
103 椭圆曲线公钥密码导引 2006.10 祝跃飞 张亚娟 著
104 椭圆与超椭圆曲线公钥密码的理论与实现 2006.12 王学理 裴定一 著
105 散乱数据拟合的模型、方法和理论 2007.1 吴宗敏 著
106 非线性演化方程的稳定性与分歧 2007.4 马天 汪宁宏 著
107 正规族理论及其应用 2007.4 顾永兴 庞学诚 方明亮 著

108 组合网络理论 2007.5 徐俊明 著
109 矩阵的半张量积:理论与应用 2007.5 程代展 齐洪胜 著
110 鞅与 Banach 空间几何学 2007.5 刘培德 著
111 非线性常微分方程边值问题 2007.6 葛渭高 著
112 戴维-斯特瓦尔松方程 2007.5 戴正德 蒋慕蓉 李栋龙 著
113 广义哈密顿系统理论及其应用(第二版) 2007.7 李继彬 赵晓华 刘正荣 著
114 Adams 谱序列和球面稳定同伦群 2007.7 林金坤 著
115 矩阵理论与应用 2007.8 陈公宁 著